A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7) AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID

RELATED APPLICATIONS

The present application claims priority to U.S. Provisional Patent Application Serial No. 60/091,315, filed June 30, 1998 and U.S. Provisional Patent Application Serial No. 60/111,909, filed December 10, 1998, the disclosures of which are incorporated herein by reference in their entireties.

FIELD OF THE INVENTION

The present invention is directed to a polynucleotide comprising open reading frames defining a coding region encoding a retinoblastoma binding protein (RBP-7) as well as regulatory regions located both at the 5'end and the 3'end of said coding region. The present invention also pertains to a polynucleotide carrying the natural regulation signals of the *RBP-7* gene which is useful in order to express a heterologous nucleic acid in host cells or host organisms as well as functionally active regulatory polynucleotides derived from said regulatory region. The invention also concerns polypeptides encoded by the coding region of the *RBP-7* gene. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention includes genetic markers, namely biallelic markers, that are means that may be useful for the diagnosis of diseases related to an alteration in the regulation or in the coding regions of the *RBP-7* gene and for the prognosis/diagnosis of an eventual treatment with therapeutic agents, especially agents acting on pathologies involving abnormal cell proliferation and/or abnormal cell differentiation.

BACKGROUND OF THE INVENTION

Among the genetic alterations that have been shown to represent direct or indirect causative agents of proliferative diseases, such as cancers, there may be cited mutations occurring at loci harboring genes that are called tumor suppressor genes.

Tumor suppressor genes are defined as genes involved in the control of abnormal cell proliferation and whose loss or inactivation is associated with the development of malignancy. Tumor suppressor genes encompass ortho-genes, emerogenes, flatogenes, and onco-suppressor genes.

More specifically, tumor suppressor genes are genes whose products inhibit cell growth. Mutant alleles in cancer cells have lost their normal function, and act in the cell in a recessive way in that both copies of the gene must be inactivated in order to change the cell phenotype. The tumor phenotype can be rescued by the wild-type allele, as shown by cell

fusion experiments first described by Harris and colleagues (Harris H. et al., 1969). Germline mutations of tumor suppressor genes may be transmitted and thus studied in both constitutional and tumor DNA from familial or sporadic cases. The current family of tumor suppressors include DNA-binding transcription factors (i.e. p53, WT1), transcription regulators (i.e., RB, APC) and protein kinase inhibitors (i.e. p16).

The existence of tumor suppressor genes has been particularly shown in cases of hereditary cancers. These are cancer where there is a clear pattern of inheritance, usually autosomal dominant, with a tendency for earlier age of onset than for sporadic tumors.

Tumor suppressor genes are detected in the form of inactivating mutations that are tumorigenic. The two best characterized genes of this class code for the proteins RB (Retinoblastoma protein) and p53.

Retinoblastoma is a human childhood disease, involving a tumor in the retina. It occurs both as an inheritable trait and sporadically (by somatic mutation). Retinoblastoma arises when both copies of the *RB* gene are inactivated. In the inherited form of the disease, one parental chromosome carries an alteration in this region, usually a deletion. A somatic event in retinal cells that causes the loss of the other copy of the *RB* gene causes a tumor. Forty percent of cases are hereditary, transmitted as an autosomal dominant trait with 90% penetrance. Of these cases, around 10-15% are transmitted from an affected parent, the remaining arising as de novo germ-lime mutations. In the sporadic form of the disease, the parental chromosomes are normal, and both *RB* alleles are lost by somatic events. The tumor suppressor nature of *RB* was shown by the introduction of a single copy of *RB1* into tumor cell lines lacking the gene, resulting in complete or partial suppression of the tumorigenic phenotype.

The RB protein has a regulatory role in cell proliferation, acting via transcription factors to prevent the transcriptional activation of a variety of genes, the products of which are required for the onset of DNA synthesis, the S phase of the cell cycle.

When investigating on the molecular function of *RB*, it has been found that the RB protein interacts with a variety of viral proteins, including several tumor antigens, such as SV40 T antigen, adenovirus E1A protein, human papillomavirus E7. These viral proteins have been shown to bind to RB, thereby inactivating it and allowing cell division to occur.

Thus, an important step toward defining a mechanism underlying tumor suppressor activity of the *RB* gene was the observation that the transforming products of adenovirus (E1A protein), simian virus 40 (large T antigen) and human papillomavirus (E7 protein) could precipitate wild-type RB protein. This, in turn, led to the identification of a family of cellular proteins that can reversibly bind to a discrete domain on the RB protein, referred to as the T/E1A pocket by using the same specificity as the viral products. The subsequent observation

that protein binding was inhibited following RB protein phosphorylation in the late G_1 phase of the cell cycle suggested the hypothesis that the RB protein, as well as the related product p107, may regulate the functional activity of its binding partners by a cell-cycle dependent pattern of physical association. In particular, the activity of the RB protein has been shown to be regulated through cell cycle-dependent phosphorylation by cyclin-dependent kinases.

The picture of transcription regulation is made even more complex by the finding that a number of RB related proteins (e.g. p107 and p130) also bind members of the E2F family and are therefore involved in regulatory process.

In view of the foregoing, there clearly exists a pressing need to identify and characterize the cellular proteins that interact with the retinoblastoma protein in order to provide diagnostic and therapeutic tools useful to prevent and cure cell differentiation disorders, particularly disorders in which a lack of completion of cell differentiation , particularly in terminal cell differentiation, or in which an abnormal cell proliferation is detected, such as in proliferative diseases like cancer.

For the purpose of the present invention, cells with abnormal proliferation include, but are not limited to, cells characteristic of the following disease states: thyroid hyperplasia, psoriasis, benign prostatic hypertrophy, cancers including breast cancer, sarcomas and other neoplasms, bladder cancer, colon cancer, lung cancer, prostate cancer, various leukemias and lymphomas.

SUMMARY OF THE INVENTION

This invention is based on the discovery of a nucleic acid molecule encoding a novel protein, more particularly a retinoblastoma binding protein (RBP-7).

The present invention pertains to nucleic acid molecules comprising the genomic sequence of the gene encoding RBP-7. The *RBP-7* genomic sequence comprises regulatory sequence located upstream (5'-end) and downstream (3'-end) of the transcribed portion of said gene, these regulatory sequences being also part of the invention.

The invention also deals with the complete cDNA sequence encoding the RBP-7 protein, as well as with the corresponding translation product.

Oligonucleotide probes or primers hybridizing specifically with a *RBP-7* genomic or cDNA sequence are also part of the present invention, as well as DNA amplification and detection methods using said primers and probes.

A further aspect of the invention is recombinant vectors comprising any of the nucleic acid sequences described above, and in particular of recombinant vectors comprising a *RBP-7* regulatory sequence or a sequence encoding a RBP-7 protein, as well as of cell hosts and transgenic non human animals comprising said nucleic acid sequences or recombinant vectors.

Finally, the invention is directed to methods for the screening of substances or molecules that inhibit the expression of *RBP-7*, as well as with methods for the screening of substances or molecules that interact with a RBP-7 polypeptide or that modulate the activity of a RBP-7 polypeptide.

The invention also concerns biallelic markers of the *RBP-7* gene which can be useful for genetic studies, for diagnosis of diseases related to an alteration in the regulation or in the coding regions of the *RBP-7* gene and for the prognosis/diagnosis of an eventual treatment with therapeutic agents, especially agents acting on pathologies involving abnormal cell proliferation and/or abnormal cell differentiation

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram showing a map of the *RBP-7* gene.

Figure 2 is a presentation of the *RBP-7* gene structure with the amplified fragments and the biallelic markers of the present invention.

BRIEF DESCRIPTION OF THE SEQUENCES PROVIDED IN THE SEQUENCE LISTING

SEQ ID No. 1 contains a genomic sequence of *RBP-7* comprising the 5' regulatory region (upstream untranscribed region), the exons and introns, and the 3' regulatory region (downstream untranscribed region).

SEQ ID No. 2 contains the 5'-regulatory sequence (upstream untrancribed region) of *RBP-7*.

SEQ ID No. 3 contains the 3'-regulatory sequence (upstream untrancribed region) of *RBP-7*.

SEQ ID No. 4 contains the *RBP-7* cDNA sequence.

SEQ ID Nos 5 to 28 contain the exons 1 to 24 of *RBP-7*.

SEQ ID No. 29 contains the protein sequence encoded by the nucleotide sequence of SEQ ID No. 4.

SEQ ID Nos 30 to 50 contain the fragments containing a polymorphic base of a biallelic marker (first allele).

SEQ ID Nos 51 to 71 contain the fragments containing a polymorphic base of a biallelic marker (second allele).

SEQ ID Nos 72 to 101 contain the amplification primers.

SEQ ID Nos 102 to 136 contain the microsequencing primers.

SEQ ID Nos 137 and 138 contain cDNA amplification primers.

SEQ ID Nos 139 and 140 respectively contain a primer containing the additional PU 5' sequence and the additional RP 5' sequence described further in Example 3.

In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to indicate the locations of biallelic markers within the sequences and to identify each of the alleles present at the polymorphic base. The code "r" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is an adenine. The code "y" in the sequences indicates that one allele of the polymorphic base is a thymine, while the other allele is a cytosine. The code "m" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an cytosine. The code "k" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a thymine. The code "s" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a cytosine. The code "w" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an thymine. The nucleotide code of the original allele for each biallelic marker is the following:

Biallelic marker	Original allele
5-124-273	A
5-127-261	C
5-130-257	A
5-130-276	A
5-131-395	A
5-135-357	A
5-136-174	T
5-140-120	T
5-143-101	C
5-143-84	G
5-145-24	A
5-148-352	T
99-1437-325	A
99-1442-224	T

In some instances, the polymorphic bases of the biallelic markers alter the identity of an amino acids in the encoded polypeptide. This is indicated in the accompanying Sequence Listing by use of the feature VARIANT, placement of an Xaa at the position of the polymorphic amino acid, and definition of Xaa as the two alternative amino acids. For example if one allele of a biallelic marker is the codon CAC, which encodes histidine, while the other allele of the biallelic marker is CAA, which encodes glutamine, the Sequence Listing for the encoded

polypeptide will contain an Xaa at the location of the polymorphic amino acid. In this instance, Xaa would be defined as being histidine or glutamine.

In other instances, Xaa may indicate an amino acid whose identity is unknown. In this instance, the feature UNSURE is used, placement of an Xaa at the position of the unknown amino acid and definition of Xaa as being any of the 20 amino acids or being unknown.

DETAILED DESCRIPTION OF THE INVENTION

The aim of the present invention is to provide polynucleotides and polypeptides related to the *RBP-7* gene and to a RBP-7 protein, which is potentially involved in the regulation of the differentiation of various cell types in mammals. A deregulation or an alteration of this protein may be involved in the generation of a pathological state in a patient. Such pathological state includes disorders caused by cell apoptosis or in contrast by an abnormal cell proliferation such as in cancers.

The unphosphorylated form of the Retinoblastoma (RB) protein specifically binds several proteins, and these interactions occur only during part of the cell cycle, prior to the S phase. The target proteins of the RB protein include E2F transcription factors and cyclins of the D and E types. Binding to the RB protein inhibits the ability of E2F to activate transcription, which suggests that the RB protein may repress the expression of genes dependent on E2F. Interaction of the RB protein with E2F-1, a member of the E2F transcription factors family, inhibits transcription of genes involved in DNA synthesis and therefore suppresses cell growth. Additionally, it has been found that the complexes formed between E2F and the RB protein are disrupted in the presence of the viral oncoproteins that bind to the RB protein, suggesting a key role of the RB protein in the regulation of E2F activity.

It has been shown that the RB protein forms two types of complexes with E2F. One of these two types involves a binary complex of the RB protein and E2F that does not bind DNA in a gel retardation assay, and the second type of RB protein/E2F complex involves another factor, RBP60, which allows the RB protein/E2F complex to bind DNA and produce a distinct complex in a gel retardation assay. One hypothesis is that RB protein might be regulating the DNA-binding as well as the transcription activation function of E2F. It has also been demonstrated that E2F can bind DNA as an oligomeric complex composed of at least two distinct proteins.

Recent reports indicate that approximately 10 proteins have been identified that bind to the RB protein using the same binding surface as the viral oncoproteins. Several of these cellular proteins, including the E2F transcription factor described above, comprise members of the *myc* oncogene family, a p46 protein (Rb-AP46), MyoD, Elf-1, protein phosphatase type 1

catalytic subunit and several proteins designated generically as "Retinoblastoma Binding Proteins" (RBBP), some of these latter proteins being defined as E2F-like proteins.

Defeo-Jones et al. (1991) have cloned the cDNA of two members of the RBBP family, namely RBP-1 and RBP-2. RBP-1 and RBP-2 bind specifically to the RB protein *in vitro*. RBP-2 has been shown to interact noncovalently with RB ptotein via the binding of a consensus amino acid sequence of RBP-2, namely the LXCXE amino acid sequence, to the conserved T/E1A pocket of the RB protein (Kim et al., 1994). This LXCXE consensus amino acid sequence is also present within the adenovirus E1A protein, the SV40 large T antigen as well as within the human papillomavirus E7 protein. RBP-1 and RBP-2 have been hypothesized to function as transcription factors, like E2F. Helin et al. (1992) have cloned a cDNA encoding another member of the RBBP family, namely RBP-3. Sakai et al. (1995) have cloned a novel RBBP protein designated as RBP-6, the locus of which has been mapped on chromosome 16 between p11.2 and p12.

For the E2F family, replicating and differentiating cells need the RB protein or RB protein family members (e.g. p107 or p130) to counterbalance its apoptotic effect. E2F induces apoptosis when over-expressed in cells with the wild type p53 gene, but favors proliferation in p53 -/- cells. E2F-induced apoptosis follows entry of the cell into S-phase. The E2F death-promoting effect can be blocked by co-expression of p105, a RB protein family member. Conversely, by gene knock-out studies, it has been demonstrated that E2F is critical for the normal development of diverse cell types. Mice null for the E2F1 gene show defects at a young age in the terminal differentiation of cell types in which apoptosis play an important role, namely T-cells or epithelial cells of the testis or of other exocrine glands. With increasing age, these animals develop wide-spread tumors. This data indicates that E2F plays a physiological role in normal development, probably by inducing apoptosis in a specific set of developing cells.

The retinoblastoma binding proteins of the E2F type have also been described in PCT Application No. WO 65/24223, PCT Application No. WO 96/25494 and in US Patent No. 5,650,287, the disclosures of which are incorporated herein by reference in their entireties. Other retinoblastoma binding proteins have been described, notably in PCT Application No. WO 94/12521 , in PCT Application No. WO 95/17198, in PCT Application No. 93/23539 and in PCT Application No. WO 93/06168, the disclosures of which are incorporated herein by reference in their entireties.

DEFINITIONS

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

The term "*RBP-7* gene", when used herein, encompasses mRNA and cDNA sequences encoding the RBP-7 protein. In the case of a genomic sequence, the *RBP-7* gene also includes native regulatory regions which control the expression of the coding sequence of the *RBP-7* gene.

The term "functionally active fragment" of the RBP-7 protein is intended to designate a polypeptide carrying at least one of the structural features of the RBP-7 protein involved in at least one of the biological functions and/or activity of the RBP-7 protein. Particularly preferred are peptide fragments carrying either the retinoblastoma protein binding domain and/or the DNA binding domain of the RBP-7 protein.

A "heterologous" or "exogenous" polynucleotide designates a purified or isolated nucleic acid that has been placed, by genetic engineering techniques, in the environment of unrelated nucleotide sequences, such as the final polynucleotide construct does not occur naturally. An illustrative, but not limitatitive, embodiment of such a polynucleotide construct may be represented by a polynucleotide comprising (1) a regulatory polynucleotide derived from the RBP-7 gene sequence and (2) a polynucleotide encoding a cytokine, for example GM-CSF. The polypeptide encoded by the heterologous polynucleotide will be termed an heterologous polypeptide for the purpose of the present invention.

By a "biologically active fragment or variant" of a regulatory polynucleotide according to the present invention is intended a polynucleotide comprising or alternatively consisting of a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host.

For the purpose of the invention, a nucleic acid or polynucleotide is "functional" as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and such sequences are "operatively linked" to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide. An operable linkage is a linkage in which the regulatory nucleic acid and the DNA sequence sought to be expressed are linked in such a way as to permit gene expression.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. More precisely, two DNA molecules (such as a polynucleotide containing a promoter region and a polynucleotide encoding a desired polypeptide or polynucleotide) are said to be "operably linked" if the nature of the linkage between the two polynucleotides does not (1) result in the introduction of a frame-shift mutation or (2) interfere with the ability of the polynucleotide containing the

promoter to direct the transcription of the coding polynucleotide. The promoter polynucleotide would be operably linked to a polynucleotide encoding a desired polypeptide or a desired polynucleotide if the promoter is capable of effecting transcription of the polynucleotide of interest.

An "altered copy" of the *RBP-7* gene is intended to designate a *RBP-7* gene that has undergone at least one substitution, addition or deletion of one or several nucleotides, wherein said nucleotide substitution, addition or deletion preferably causes a change in the amino acid sequence of the resulting translation product or alternatively causes an increase or a decrease in the expression of the *RPB-7* gene.

The terms "sample" or "material sample" are used herein to designate a solid or a liquid material suspected to contain a polynucleotide or a polypeptide of the invention. A solid material may be, for example, a tissue slice or biopsy which is searched for the presence of a polynucleotide encoding a RBP-7 protein, either a DNA or RNA molecule or within which is searched for the presence of a native or a mutated RBP-7 protein, or alternatively the presence of a desired protein of interest the expression of which has been placed under the control of a *RBP-7* regulatory polynucleotide. A liquid material may be, for example, any body fluid like serum, urine etc., or a liquid solution resulting from the extraction of nucleic acid or protein material of interest from a cell suspension or from cells in a tissue slice or biopsy. The term "biological sample" is also used and is more precisely defined within the Section dealing with DNA extraction.

As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification if starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. As an example, purification from 0.1% concentration to 10% concentration is two orders of magnitude.

The term "isolated" requires that the material be removed from its original environment (e.g. the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition and still be isolated in that the vector or composition is not part of its natural environment.

Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or an oligonucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself

and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule.

As used interchangeably herein, the term "oligonucleotides", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. However, the polynucleotides of the invention are preferably comprised of greater than 50% conventional deoxyribose nucleotides, and most preferably greater than 90% conventional deoxyribose nucleotides. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population which are heterozygous at a particular allele. In a biallelic system, the heterozygosity rate is on average equal to $2P_a(1-P_a)$, where P_a is the frequency of the least common allele. In order to be useful in genetic studies, a genetic marker should have an adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker consists of determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation

occurs. A single nucleotide polymorphism is a single base pair change. Typically a single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide, also give rise to single nucleotide polymorphisms. In the context of the present invention "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. However, the polymorphism can also involve an insertion or a deletion of at least one nucleotide, preferably between 1 and 5 nucleotides. The nucleotide modification can also involve the presence of several adjacent single base polymorphisms. This type of nucleotide modification is usually called a "variable motif". Generally, a "variable motif" involves the presence of 2 to 10 adjacent single base polymorphisms. In some instances, series of two or more single base polymorphisms can be interrupted by single bases which are not polymorphic. This is also globally considered to be a "variable motif". Typically, between different genomes or between different individuals, the polymorphic site may be occupied by two different nucleotides.

The term "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a single nucleotide polymorphism having two alleles at a fairly high frequency in the population. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site. Typically, the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker".

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the

polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on.

As used herein the terminology "<u>defining a biallelic marker</u>" means that a sequence includes a polymorphic base from a biallelic marker. The sequences defining a biallelic marker may be of any length consistent with their intended use, provided that they contain a polymorphic base from a biallelic marker. The sequence is preferably between 1 and 500 nucleotides in length, more preferably between 5, 10 , 15, 20, 25, or 40 and 200 nucleotides and still more preferably between 30 and 50 nucleotides in length. Each biallelic marker therefore corresponds to two forms of a polynucleotide sequence included in a gene, which, when compared with one another, present a nucleotide modification at one position. Preferably, the sequences defining a biallelic marker include a polymorphic base selected from the group consisting of biallelic markers A1 to A21. In some embodiments the sequences defining a biallelic marker comprise one of the sequences selected from the group consisting of SEQ ID Nos 30 to 71. Likewise, the term "marker" or "biallelic marker" requires that the sequence is of sufficient length to practically (although not necessarily unambiguously) identify the polymorphic allele, which usually implies a length of at least 4, 5, 6, 10, 15, 20, 25, or 40 nucleotides.

Variants And Fragments

<u>1. Polynucleotides</u>

The invention also relates to variants and fragments of the polynucleotides described herein, particularly of a *RBP-7* gene containing one or more biallelic markers according to the invention.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences that are at least 95% identical to any of SEQ ID Nos 1-28 or the sequences complementary thereto or to any polynucleotide fragment of at least 8 consecutive nucleotides of any of SEQ ID Nos 1-28 or the sequences complementary thereto, and preferably at least 98% identical, more particularly at least 99.5% identical, and most preferably at least 99.9% identical to any of SEQ ID Nos 1-28 or the sequences complementary thereto or to any polynucleotide fragment of at least 8 consecutive nucleotides of any of SEQ ID Nos 1-28 or the sequences complementary thereto.

Changes in the nucleotide of a variant may be silent, which means that they do not alter the amino acids encoded by the polynucleotide.

However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

In the context of the present invention, particularly preferred embodiments are those in which the polynucleotides encode polypeptides which retain substantially the same biological function or activity as the mature RBP-7 protein.

A polynucleotide fragment is a polynucleotide having a sequence that entirely is the same as part but not all of a given nucleotide sequence, preferably the nucleotide sequence of a *RBP-7* gene, and variants thereof. The fragment can be a portion of an exon or of an intron of a *RBP-7* gene. It can also be a portion of the regulatory sequences of the *RBP-7* gene. Preferably, such fragments comprise the polymorphic base of at least one of the biallelic markers of SEQ ID Nos. 30-71.

Such fragments may be "free-standing", i.e. not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region. However, several fragments may be comprised within a single larger polynucleotide.

As representative examples of polynucleotide fragments of the invention, there may be mentioned those which are from about 4, 6, 8, 15, 20, 25, 40, 10 to 20, 10 to 30, 30 to 55, 50 to 100, 75 to 100 or 100 to 200 nucleotides in length. Preferred are those fragments which are about 47 nucleotides in length, such as those of SEQ ID Nos 30-71 or the sequences complementary thereto and containing at least one of the biallelic markers of a *RBP-7* gene which are described herein. It will of course be understood that the polynucleotides of SEQ ID Nos 30-71 or the sequences complementary thereto can be shorter or longer, although it is

preferred that they at least contain the polymorphic base of the biallelic marker which can be located at one end of the fragment or in the internal portion of the fragment.

2. Polypeptides.

The invention also relates to variants, fragments, analogs and derivatives of the polypeptides described herein, including mutated RBP-7 proteins.

The variant may be 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated RBP-7 is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mutated RBP-7, such as a leader or secretory sequence or a sequence which is employed for purification of the mutated RBP-7 or a preprotein sequence. Such variants are deemed to be within the scope of those skilled in the art.

More particularly, a variant RBP-7 polypeptide comprises amino acid changes ranging from 1, 2, 3, 4, 5, 10 to 20 substitutions, additions or deletions of one amino acid, preferably from 1 to 10, more preferably from 1 to 5 and most preferably from 1 to 3 substitutions, additions or deletions of one amino acid. The preferred amino acid changes are those which have little or no influence on the biological activity or the capacity of the variant RBP-7 polypeptide to be recognized by antibodies raised against a native RBP-7 protein.

As illustrative embodiments of variant RBP-7 polypeptides encompassed by the present invention, there are the following polypeptides :

- a polypeptide comprising a Glycine residue at the amino acid position 293 of the amino acid sequence of SEQ ID No. 29;
- a polypeptide comprising a Glutamic acid at the amino acid in position 963 of SEQ ID No. 29; and,
- a polypeptide comprising a Methionine residue at the amino acid position 969 of the amino acid sequence of SEQ ID No. 29.

By homologous peptide according to the present invention is meant a polypeptide containing one or several amino acid additions, deletions and/or substitutions in the amino acid sequence of a RBP-7 polypeptide. In the case of an amino acid substitution, one or several - consecutive or non-consecutive- amino acids are replaced by "equivalent" amino acids. The expression "equivalent" amino acid is used herein to designate any amino acid that may substituted for one of the amino acids belonging to the native protein structure without decreasing the binding properties of the corresponding peptides to the retinoblastoma proteins

(i.e. RBP, p130, p107 etc.). In other words, the "equivalent" amino acids are those which allow the generation or the synthesis of a polypeptide with a modified sequence when compared to the amino acid sequence of the native RBP-7 protein, said modified polypeptide being able to bind to the retinoblastoma protein and/or to induce antibodies recognizing the parent polypeptide comprising, consisting essentially of, or consisting of a RBP-7 polypeptide.

These equivalent amino acids may be determined either by their structural homology with the initial amino acids to be replaced, by the similarity of their net charge, and optionally by the results of the cross-immunogenicity between the parent peptides and their modified counterparts.

By an equivalent amino acid according to the present invention is also meant the replacement of a residue in the L-form by a residue in the D form or the replacement of a Glutamic acid (E) residue by a Pyro-glutamic acid compound. The synthesis of peptides containing at least one residue in the D-form is, for example, described by Koch (Koch Y., 1977, Biochem. Biophys. Res. Commun., Vol.74:488-491).

A specific, but not restrictive, embodiment of a modified peptide molecule of interest according to the present invention, which comprises, consists essentially of, or consists of a peptide molecule which is resistant to proteolysis, is a peptide in which the -CONH- peptide bond is modified and replaced by a (CH_2NH) reduced bond, a (NHCO) retro inverso bond, a (CH_2-O) methylene-oxy bond, a (CH_2-S) thiomethylene bond, a (CH_2CH_2) carba bond, a (CO-CH_2) cetomethylene bond, a (CHOH-CH_2) hydroxyethylene bond), a (N-N) bound, a E-alcene bond or also a -CH=CH- bond.

A polypeptide fragment is a polypeptide having a sequence that entirely is the same as part but not all of a given polypeptide sequence, preferably a polypeptide encoded by a *RBP-7* gene and variants thereof. Preferred fragments include those regions possessing antigenic properties and which can be used to raise antibodies against the RBP-7 protein.

Such fragments may be "free-standing", i.e. not part of or fused to other polypeptides, or they may be comprised within a single larger polypeptide of which they form a part or region. However, several fragments may be comprised within a single larger polypeptide.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which comprise at least about 5, 6, 7, 8, 9 or 10 to 15, 10 to 20, 15 to 40, or 30 to 55 amino acids of the RBP-7 protein. In some embodiments, the fragments contain at least one amino acid mutation in the RBP-7 protein.

Complementary Polynucleotides

For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired

with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G.

Identity Between Nucleic Acids Or Polypeptides

The terms "percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988; Altschul et al., 1990; Thompson et al., 1994; Higgins et al., 1996; Altschul et al., 1990; Altschul et al., 1993). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990; Altschul et al., 1990, 1993, 1997). In particular, five specific BLAST programs are used to perform the following task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by

means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., 1992; Henikoff and Henikoff, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990). The programs listed above may be used with the default parameters or with modified parameters provided by the user.

RBP-7 GENE, CORRESPONDING CDNAS AND *RBP-7* CODING AND REGULATORY SEQUENCES

The gene encoding a RBP-7 polypeptide has been found by the inventors to be located on human chromosome 1, more precisely within the 1q43 locus of said chromosome. The *RBP-7* gene has a length of about 166 kilobases and contains a 5' regulatory region, 24 exons, and a 3' regulatory region. A 5'-UTR region is spans the whole Exon 1 and the major portion of the 5' end of Exon 2. A 3'-UTR region is spans the major portion of the 3' end of Exon 24.

The present invention first concerns a purified or isolated nucleic acid encoding a Retinoblastoma Binding Protein named RBP-7 as well as a nucleic acid complementary thereto and fragments and variants thereof.

In particular, the invention concerns a purified or isolated nucleic acid comprising at least 8 consecutive nucleotides of a polynucleotide selected from the group consisting of SEQ ID Nos 1 and 4 as well as a nucleic acid sequence complementary thereto and fragments and variants thereof. The length of the fragments described above can range from at least 8, 10, 15, 20 or 30 to 200 nucleotides, preferably from at least 10 to 50 nucleotides, more preferably from at least 40 to 50 nucleotides. In some embodiments, the fragments may comprise more than 200 nucleotides of SEQ ID Nos. 1 and 4 or the sequences complementary thereto.

The invention also pertains to a purified or isolated nucleic acid of at least 8 nucleotides in length that hybridizes under stringent hybridization conditions with a polynucleotide selected from the group consisting of SEQ ID Nos 1 and 4 or the sequences complementary thereto. The length of the nucleic acids described above can range from 8, 10, 15, 20 or 30 to 200 nucleotides, preferably from 10 to 50 nucleotides, more preferably from 40 to 50 nucleotides. Such nucleic acids may be used as probes or primers, such as described in the corresponding section of the present specification.

The invention also encompasses a purified, isolated, or recombinant polynucleotide comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity

with a nucleotide sequence of SEQ ID Nos 1 and 4 or a complementary sequence thereto or a fragment thereof. Percent identity may be determined using any of the programs and scoring matrices described above. For example, percent identity may be determined using BLASTN with the default parameters. In addition, the scoring matrix may be BLOSUM62

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No. 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No. 1: 1-481, 666-1465, 1521-67592, 67704-71118, 71185-72598, 72690-75543, 75624-81841, 81934-83019, 83406-87901, 88041-93856, 93937-97158, 97236-98962, 99086-103188, 103745-104303, 104654-105084, 105180-106682, 106781-107798, 107897-108392, 108552-114335, 114418-114491, 114594-132246, 132332-134150, 134350-145565, 145842-146332, 146775-150446, 150542-152959, 153176-155590, 155738-159701, 160466-161028, 161453-162450. Additional preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No. 4 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No. 4: 1-208, 1307-1350, 1703-1865, 2107-2180, 2843-3333, 3871-3882, 4222-4276, and 5017-5579. It should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section.

The main structural features of the *RBP-7* gene are shown in Figure 1. The upper line shows a structural map of the polynucleotide of SEQ ID No. 1 including the 24 exons, that are indicated by closed boxes, and the 23 introns, as well the 5'- and 3'-flanking regulatory regions. The position of the first nucleotide at 5'end of each exon is also indicated, the nucleotide at position 1 being the first nucleotide at the 5' end of the polynucleotide of SEQ ID No. 1.

Generally, an intron is defined as a nucleotide sequence that is present both in the genomic DNA and in the unspliced mRNA molecule, and which is absent from the mRNA molecule which has already gone through splicing events.

For the purpose of the present invention and in order to make a clear and unambiguous designation of the different nucleic acids encompassed, it has been postulated that the polynucleotides contained both in the nucleotide sequence of SEQ ID No. 1 and in the nucleotide sequences of SEQ ID No. 4 are considered as exonic sequences. Conversely, the polynucleotides contained in the nucleotide sequence of SEQ ID No. 1 and located between

Exon 1 and Exon 24, but which are absent both from the nucleotide sequence of SEQ ID No. 4 are considered as intronic sequences.

More precisely, the structural characteristics of the *RBP-7* gene, as represented in Figure 1 are as follows :

a) a regulatory region, located between the nucleotide at position 1 and the nucleotide at position 273 of SEQ ID No. 1;

b) a "coding" region, located between the nucleotide at position 274 and the nucleotide at position 161451 of SEQ ID No. 1, comprising 24 exons and 23 introns, wherein said region defines the *RBP-7* coding region.

c) a regulatory region, beginning at the nucleotide at position 161452 and ending at the nucleotide in position 162450 (the 3'-end nucleotide) of SEQ ID No. 1.

The translation start site ATG is located within the second exon and the translation stop codon is located within Exon 24 of the nucleotide sequence of SEQ ID No. 1.

The middle line of Figure 1 shows the cDNA corresponding to the longest *RBP-7* mRNA including the 24 exons. Each exon is represented by a specific box. The numbers located under the exon boxes indicate the nucleotide position of the 5'end polynucleotide of each exon, it being understood that the nucleotide at position 1 is the 5'end nucleotide of the cDNA. pAd denotes the four potential polyadenylation sites.

The lower line of Figure 1 shows a map of the *RBP-7* coding sequence (CDS), the start codon being located from the nucleotide in position 442 to the nucleotide in position 444 of the *RBP-7* cDNA of SEQ ID No. 4 and the stop codon being located from the nucleotide in position 4378 to the nucleotide in position 4380 of the *RBP-7* cDNA of SEQ ID No. 4.

The 24 exons included in the *RBP-7* gene are represented in Figure 1 and are described in Table A.

TABLE A

Exon	SEQ ID No.	Beginning position in SEQ ID No. 1	End position In SEQ ID No. 1
1	5	274	665
2	6	1466	1520
3	7	67593	67703
4	8	71119	71184
5	9	72599	72689
6	10	75544	75623
7	11	81842	81933
8	12	87902	88040
9	13	93857	93936
10	14	97159	97235
11	15	98963	99117

Exon	SEQ ID No.	Begining position in SEQ ID No. 1	End position In SEQ ID No. 1
12	16	103570	103642
13	17	105085	105179
14	18	106683	106780
15	19	107799	108042
16	20	108376	108551
17	21	114336	114593
18	22	132247	132331
19	23	134151	134349
20	24	145566	146774
21	25	150447	150560
22	26	152960	153175
23	27	155591	155737
24	28	159702	161451

The middle line depicts the main structural features of a purified or isolated nucleic acid consisting of the longest cDNA that is obtained after reverse transcribing a mRNA generated after transcription of the *RBP-7* gene. The longest mRNA has a nucleotide length of about 6 kilobases.

As it is depicted in Figure 1, the main characteristics of the longest *RBP-7* cDNA are the following :

a) A 5'-UTR region extending from the nucleotide at position 1 to the nucleotide at position 441 of SEQ ID No. 4;

b) An open reading frame (ORF) encoding the longest form of RBP-7 protein, wherein said ORF extends from the nucleotide at position 442 to the nucleotide at position 4380 of SEQ ID No. 4. The ATG translation start site is located between the nucleotide at position 442 and the nucleotide at position 444 of SEQ ID No. 4. The stop codon is located between the nucleotide at position 4378 and the nucleotide at position 4380 of SEQ ID No. 4.

c) A 3'-UTR region extending from the nucleotide at position 4381 to the nucleotide at position 6002 of SEQ ID No. 4. This 3'-UTR region contains four potential polyadenylation sites comprising respectively the nucleotides between positions 4878 and 4883, 5116 and 5121, 5896 and 5901 and between positions 5981 and 5986 of SEQ ID No. 4.

Figure 2 is a representation of the *RBP-7* gene in which the 24 exons are shown as closed boxes.

a) In each closed box that represents a given Exon, there are indicated both a number of base pairs corresponding to the non coding sequence eventually present in this Exon, and a number of amino acids. The number of amino acids is calculated as follows, starting from Exon 2 : Exon 2 contains two complete codons and the first base of a third codon; only the two

complete codons are taken into account and the additional base is taken into account as the first base of the first codon of Exon 3, etc.;

b) The arrows above the Intron lines or above the Exon boxes indicate the localization of the different polymorphic markers of the invention on the *RBP-7* gene, as well as their marker names;

c) The bold letters above exons 11 and 20 indicate the effect of the base changes constitutive to these polymorphic markers on the amino acid sequence of the resulting RBP-7 translation product.

The polynucleotide of SEQ ID No. 4 contains, from its 5' end to its 3' end, the sequences resulting from the 24 exons located in Table A on the *RBP-7* genomic sequence, said exonic sequences being positioned on the *RBP-7* cDNA of SEQ ID No. 4, as detailed in Table B below.

TABLE B

Exon	SEQ ID No.	Beginning position in SEQ ID No. 4	End position In SEQ ID No. 4
1	5	1	392
2	6	393	447
3	7	448	558
4	8	559	624
5	9	625	715
6	10	716	795
7	11	796	887
8	12	888	1026
9	13	1027	1106
10	14	1107	1183
11	15	1184	1338
12	16	1339	1411
13	17	1412	1507
14	18	1508	1604
15	19	1605	1848
16	20	1849	2024
17	21	2025	2282
18	22	2283	2367
19	23	2368	2566
20	24	2567	3775
21	25	3776	3889
22	26	3890	4105
23	27	4106	4252
24	28	4253	6002

The nucleotide sequence of the *RBP-7* cDNA possesses some homologies with a cDNA encoding another human retinoblastoma binding protein, namely hRBP-1. This homology is

randomly distributed throughout the whole cDNA sequences, without visible nucleic acid regions that are characteristic of conserved regions between cDNA sequences encoding different retinobastoma binding proteins.

The majority of interrupted genes are transcribed into a RNA that gives rise to a single type of spliced mRNA. But the RNAs of some genes follow patterns of alternative splicing, wherein a single gene gives rise to more than one mRNA species. In some cases, the ultimate pattern of expression is dictated by the primary transcript, because the use of different startpoints or termination sequences alters the splicing pattern. In other cases, a single primary transcript is spliced in more than one way, and internal exons are substituted, added or deleted. In some cases, the multiple products all are made in the same cell, but in others, the process is regulated so that particular splicing patterns occur only under particular conditions.

In the case of retinoblastoma binding proteins, alternative splicing patterns have been observed during the processing of the *RBP1* pre-mRNA (Otterson et al., 1993). More precisely, alternative splicing of *RBP1* clusters has been observed within a 207-nucleotide internal exon. From the four forms of mRNA detected, three of the predicted RBP1 peptides share amino-terminal and carboxy-terminal domains, while a fourth species encodes a distinct carboxy-terminal domain. Functional analysis of these peptides demonstrated that they are capable of precipitating retinoblastoma protein *in vitro* from K562 cell lysates, but cannot bind to mutant RB protein.

The inventors have found that a mRNA of about 6 kilobases and containing exon 1 of the *RBP-7* gene at its 5'end and exon 24 of the *RBP-7* gene at its 3' end, is produced in isolated cells from the prostate tissue, as described in Example 1.

Because the *RBP-7* gene contains a large number of exons, it is expected that the corresponding pre-mRNA is processed in a family of mRNA molecules as a result of multiple alternative splicing events.

Additionally, individually combining each polynucleotide molecule defining a specific exon of the *RBP-7* gene with at least one polynucleotide molecule defining another exon of the *RBP-7* gene will give rise to a family of translation products that may be assayed for their biological functions of interaction with retinoblastoma proteins (i.e. pRb, p107, p130 etc.) or of interaction with DNA sequences of the type recognized by the transcription factors of the E2F family. Such translation products have a shorter size than that of the resulting protein encoded by the longest *RBP-7* mRNA and thus may be advantageously used in therapeutics, as compared with the longest polypeptides, due to their weaker immunogenicity, for example.

Consequently, a further aspect of the present invention is a purified or isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 5-28 or the sequences complementary thereto.

The invention also deals with a purified or isolated nucleic acid comprising a combination of at least two polynucleotides selected from the group consisting of SEQ ID Nos 5-28 or the sequences complementary thereto, wherein the polynucleotides are ordered within the nucleic acid, from the 5' end to the 3' end of said nucleic acid, in the same order as in the SEQ ID No. 1.

In this specific embodiment of a purified or isolated nucleic acid according to the invention, said nucleic acid preferably comprises SEQ ID Nos 5 and 6 at its 5' end and SEQ ID No. 28 at its 3' end.

Regulatory Regions

As already mentioned hereinbefore, the polynucleotide of SEQ ID No. 1 contains regulatory regions both in the non-coding 5'-flanking region (SEQ ID No. 2) and the non-coding 3'-flanking region (SEQ ID No. 3) that border the coding sequences.

The promoter activity of the regulatory region contained in SEQ ID No. 1 can be assessed as described below.

Genomic sequences lying upstream of the *RBP-7* gene are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, β galactosidase, or green fluorescent protein. The sequences upstream of the *RBP-7* coding region are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for increasing transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Promoter sequences within the upstream genomic DNA may be further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter

reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter, individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

Polynucleotides carrying the regulatory elements located both at the 5' end and at the 3' end of the *RBP-7* coding region may be advantageously used to control the transcriptional and translational activity of an heterologous polynucleotide of interest.

A 5' regulatory polynucleotide of the invention may include the 5'-untranslated region (5'-UTR) or the sequence complementary thereto, or a biologically active fragment or variant thereof. The 5'-regulatory polynucleotide harbors a CAAT box from the nucleotide in position 139 to the nucleotide in position 147 of the nucleotide sequence of SEQ ID No. 2. Additionally, the 5'-regulatory polynuceotide of the invention comprises a TATA box from the nucleotide in position 199 to the nucleotide in position 205 of the nucleotide sequence of SEQ ID No. 2.

A 3' regulatory polynucleotide of the invention may include the 3'-untranslated region (3'-UTR) or the sequences complementary thereto, or a biologically active fragment or variant thereof.

Another aspect of the present invention is a purified and/or isolated polynucleotide located at the 5'end of the start codon of the *RBP-7* gene, wherein said polynucleotide carries expression and/or regulation signals allowing the expression of the *RBP-7* gene. Thus, another part of the present invention is a purified or isolated nucleic acid comprising a nucleotide sequence of SEQ ID No. 2 and functionally active fragments or variants thereof. The fragments may be of any length to facilitate the expression and/or regulation of a gene operably linked thereto. In particular, the fragments may contain one or more binding sites for transcription factors. In some embodiments, the fragments at least 8, 10, 15, 20 or 30 to 200 nucleotides of SEQ ID No. 2. In other embodiments, the fragments may comprise more than 200 nucleotides of SEQ ID No. 2 or the sequence complementary thereto.

The invention further deals with a purified and/or isolated polynucleotide located at the 3'end of the stop codon of the *RBP-7* gene, wherein said polynucleotide carries regulation signals involved in the expression of the *RBP-7* gene. Thus another part of the present invention is a purified or isolated nucleic acid comprising a nucleotide sequence of SEQ ID No. 3, the sequence complementary thereto, and functionally active fragments or variants thereof. The fragments may be of any length to facilitate the expression and/or regulation of a gene operationally linked thereto. In some embodiments, the fragments may comprise at least 8, 10,

15, 20 or 30 to 200 nucleotides of SEQ ID No. 3 or the sequence complementary thereto. In other embodiments, the fragments may comprise more than 200 nucleotides of SEQ ID No. 3 or the sequence complementary thereto.

Thus, the invention also pertains to a purified or isolated nucleic acid which is selected from the group consisting of :

a) a nucleic acid comprising the nucleotide sequence SEQ ID No. 2 or the sequence complementary thereto ;

b) a nucleic acid comprising a biologically active fragment or variant of the nucleic acid of SEQ ID No. 2 or the sequence complementary thereto.

In a specific embodiment of the above nucleic acid, said nucleic acid includes the 5'-untranslated region (5'-UTR) located between the nucleotide at position 1 to the nucleotide at position 441 of SEQ ID No. 4, or the sequences complementary thereto, or a biologically active fragment or variant thereof.

Another aspect of the present invention is a purified or isolated nucleic acid which is selected from the group consisting of :

a) a nucleic acid comprising the nucleotide sequence SEQ ID No. 3 or the sequence complementary thereto;

b) a nucleic acid comprising a biologically active fragment, a variant of the nucleic acid of SEQ ID No. 3 or the sequence complementary thereto.

In a specific embodiment of the above nucleic acid, said nucleic acid includes the 3'-untranslated region (3'-UTR) located between the nucleotide at position 4381 and the nucleotide at position 6002 of SEQ ID No. 4, or the sequences complementary thereto, or a biologically active fragment or variant thereof.

Preferred fragments of the nucleic acid of SEQ ID No. 2 or the sequence complementary thereto have a range of length from 100, 125, 150, 175, 200 to 225, 250, 273 consecutive nucleotides. Preferred fragments will comprise both the CAAT box and the TATA box of the nucleotide sequence of SEQ ID No. 2.

Preferred fragments of the nucleic acid of SEQ ID No. 3 or the sequence complementary thereto have a length of about 600 nucleotides, more particularly of about 300 nucleotides, more preferably of about 200 nucleotides and most preferably about 100 nucleotides.

In order to identify the relevant biologically active polynucleotide derivatives of SEQ ID No. 3, one may follow the procedures described in Sambrook et al. (1989, the disclosure of which is incorporated herein by reference) relating to the use of a recombinant vector carrying a marker gene (i.e. β galactosidase, chloramphenicol acetyl transferase, etc.) the expression of

which will be detected when placed under the control of a biologically active derivative polynucleotide of SEQ ID No. 3.

Regulatory polynucleotides of the invention may be prepared from the nucleotide sequence of SEQ ID No. 1 or the sequences complementary thereto by cleavage using the suitable restriction enzymes, as described in Sambrook et al. (1989), supra.

Regulatory polynucleotides may also be prepared by digestion of the nucleotide sequence of SEQ ID No. 1 or the sequences complementary thereto by an exonuclease enzyme, such as Bal31 (Wabiko et al., 1986).

These regulatory polynucleotides can also be prepared by nucleic acid chemical synthesis, as described elsewhere in the specification, when oligonucleotide probes or primers synthesis is disclosed.

The regulatory polynucleotides according to the invention may advantageously be part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism. The recombinant expression vectors according to the invention are described elsewhere in the specification.

The above defined polynucleotides that carry the expression and/or regulation signals of the *RBP-7* gene may be used, for example as part of a recombinant vector, in order to drive the expression of a desired polynucleotide, said desired polynucleotide being either (1) a polynucleotide encoding a RBP-7 protein, or a fragment or variant thereof, or (2) an "heterologous" polynucleotide, such as a polynucleotide encoding a desired "heterologous" polypeptide or a desired RNA in a recombinant cell host.

The invention also encompasses a polynucleotide comprising, consisting essentially of , or consisting of :

a) a nucleic acid comprising a regulatory polynucleotide of SEQ ID No. 2, or the sequence complementary thereto, or a biologically active fragment or variant thereof;

b) a polynucleotide encoding a desired polypeptide or nucleic acid.

c) Optionally, a nucleic acid comprising a regulatory polynucleotide of SEQ ID No. 3, or the sequence complementary thereto, or a biologically active fragment or variant thereof.

In a preferred embodiment, a polynucleotide such as disclosed above comprises the nucleic acid of SEQ ID No. 2, or the sequences complementary thereto, or a fragment, a variant or a biologically active derivative thereof which is located at the 5'end of the polynucleotide encoding the desired polypeptide or polynucleotide.

In another embodiment, a polynucleotide such as that above described comprises the nucleic acid of SEQ ID No. 3, or the sequence complementary thereto, or a fragment, a variant or a biologically active derivative thereof which is located at the 3' end of the polynucleotide

encoding the desired polypeptide or nucleic acid. A preferred desired nucleic acid comprises of a ribonucleic acid useful as antisense molecule.

The desired polypeptide encoded by the above described nucleic acid may be of various nature or origin, encompassing proteins of prokaryotic or eukaryotic origin. Among the polypeptides which may be expressed under the control of a *RBP-7* regulatory region are bacterial, fungal or viral antigens. Are also encompassed eukaryotic proteins such as intracellular proteins, such as "house keeping" proteins, membrane-bound proteins, such as receptors, and secreted proteins such as the numerous endogenous mediators including cytokines.

The desired nucleic acid encoded by the above described polynucleotide, usually a RNA molecule, may be complementary to a *RBP-7* coding sequence and thus useful as an antisense polynucleotide.

Such a polynucleotide may be included in a recombinant expression vector in order to express a desired polypeptide or a desired polynucleotide in host cell or in a host organism. Suitable recombinant vectors that contain a polynucleotide such as described hereinbefore are disclosed elsewhere in the specification.

Coding Regions

As depicted in Figure 1, the *RBP-7* open reading frame is contained in the longest *RBP-7* which mRNA has a nucleotide length of about 4 kilobases.

More precisely, the effective *RBP-7* coding sequence (CDS) is between the nucleotide at position 442 and the nucleotide at position 4377 of SEQ ID No. 4.

The invention further provides a purified or isolated nucleic acid comprising a polynucleotide selected from the group consisting of a polynucleotide comprising a nucleic acid sequence located between the nucleotide at position 442 and the nucleotide at position 4377 of SEQ ID No. 4, or the sequence complementary thereto, or a variant or fragment thereof or a sequence complementary thereto.

A further object of the present invention comprises polynucleotide fragments of the *RBP-7* gene that are useful for the detection of the presence of an unaltered or an altered copy of the *RBP-7* gene within the genome of a host organism and also for the detection and/or quantification of the expression of the *RBP-7* gene in said host organism.

Thus, another object of the present invention is a purified or isolated nucleic acid encoding a variant or a mutated RBP-7 protein.

A first preferred embodiment of a copy of the *RBP-7* gene comprises an allele in which a single base substitution in the codon encoding the Aspartic acid (D) residue in amino acid

position 293 of the RBP-7 protein of SEQ ID No. 29 leads to the amino acid replacement for a Glycine (G) residue .

A second preferred embodiment of a copy of the *RBP-7* gene comprises an allele in which a single base substitution in the codon encoding the Glycine (G) residue in amino acid position 963 of the RBP-7 protein of SEQ ID No. 29 leads to the amino acid replacement for a Glutamic acid (E) residue.

A third preferred embodiment of a copy of the *RBP-7* gene comprises an allele in which a single base substitution in the codon encoding the Leucine (L) residue in amino acid position 969 of the RBP-7 protein of SEQ ID No. 29 leads to the amino acid replacement for a Methionine (M) residue.

Thus, another object of the present invention is a purified or isolated nucleic acid encoding a mutated RBP-7 protein.

The above disclosed polynucleotide that contains only coding sequences derived from the *RBP-7* ORF may be expressed in a desired host cell or a desired host organism, when said polynucleotide is placed under the control of suitable expression signals. Such a polynucleotide, when placed under the suitable expression signals, may be inserted in a vector for its expression.

OLIGONUCLEOTIDE PROBES AND PRIMERS

Polynucleotides derived from the *RBP-7* gene described above are useful in order to detect the presence of at least a copy of a nucleotide sequence of SEQ ID No. 1, or a fragment or a variant thereof in a test sample.

The present invention concerns a purified or isolated nucleic acid comprising at least 8 consecutive nucleotides of the nucleotide sequence SEQ ID No. 1 or a sequence complementary thereto or variants thereof. In another embodiment, the present invention relates to nucleic acids comprising at least 8, 10, 15, 20 or 30 to 200 nucleotides, preferably from at least 10 to 50 nucleotides, more preferably from at least 40 to 50 nucleotides of SEQ ID No. 1 or the sequence complementary thereto. In some embodiments, the nucleic acids may comprise more than 200 nucleotides of SEQ ID No. 1 or the sequence complementary thereto.

Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No. 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No. 1: 1-481, 666-1465, 1521-67592, 67704-71118, 71185-72598, 72690-75543, 75624-81841, 81934-83019, 83406-87901, 88041-93856, 93937-97158, 97236-98962, 99086-103188, 103745-104303, 104654-105084, 105180-106682,

106781-107798, 107897-108392, 108552-114335, 114418-114491, 114594-132246, 132332-134150, 134350-145565, 145842-146332, 146775-150446, 150542-152959, 153176-155590, 155738-159701, 160466-161028, 161453-162450.

The invention also relates to an oligonucleotide of at least at least 8 nucleotides in length that hybridizes under stringent hybridization conditions with a nucleic acid selected from the group consisting of the nucleotide sequences 1-481, 666-1465, 1521-67592, 67704-71118, 71185-72598, 72690-75543, 75624-81841, 81934-83019, 83406-87901, 88041-93856, 93937-97158, 97236-98962, 99086-103188, 103745-104303, 104654-105084, 105180-106682, 106781-107798, 107897-108392, 108552-114335, 114418-114491, 114594-132246, 132332-134150, 134350-145565, 145842-146332, 146775-150446, 150542-152959, 153176-155590, 155738-159701, 160466-161028, 161453-162450 of SEQ ID No. 1 or a variant thereof or a sequence complementary thereto. In some embodiments, the invention relates to sequences comprising at least 8, 10, 15, 20 or 30 to 200 nucleotides, preferably from at least 10 to 50 nucleotides, more preferably from 40 to 50 nucleotides of SEQ ID No. 1 or the sequence complementary thereto or variants thereof. In some embodiments, the invention relates to sequences comprising more than 200 nucleotides of SEQ ID No. 1 or the sequence complementary thereto.

For the purpose of defining such a hybridizing nucleic acid according to the invention, the stringent hybridization conditions are the following :

the hybridization step is realized at 65°C in the presence of 6 x SSC buffer, 5 x Denhardt's solution, 0,5% SDS and 100μg/ml of salmon sperm DNA.

The hybridization step is followed by four washing steps :

- two 5 min washings, preferably at 65°C in a 2 x SSC and 0.1%SDS buffer;
- one 30 min washing, preferably at 65°C in a 2 x SSC and 0.1% SDS buffer,
- one 10 min washing, preferably at 65°C in a 0.1 x SSC and 0.1%SDS buffer,

the above hybridization conditions are suitable for a nucleic acid molecule of about 20 nucleotides in length. There is no need to say that the hybridization conditions described above can readily be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art. The hybridization conditions may for example be adapted according to the teachings disclosed in the book of Hames and Higgins (1985), the disclosure of which is incorporated herein by reference.

Another aspect of the invention is a purified or isolated nucleic acid comprising at least 8 consecutive nucleotides of the nucleotide sequence SEQ ID No. 4 or the sequence complementary thereto or variants thereof. In another embodiment, the nucleic acid comprises from at least 8, 10, 15, 20 or 30 to 200 nucleotides, preferably from at least 10 to 50 nucleotides,

more preferably from at least 40 to 50 nucleotides of SEQ ID No. 4 or the sequence complementary thereto or variants thereof. In some embodiments, the fragments may comprise more than 200 nucleotides of SEQ ID No. 4 or the sequence complementary thereto or variants thereof.

Additional preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No. 4 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No. 4: 1-208, 1307-1350, 1703-1865, 2107-2180, 2843-3333, 3871-3882, 4222-4276, and 5017-5579.

Alternatively, the invention also relates to an oligonucleotide of at least 8 nucleotides in length that hybridizes under the stringent hybridization conditions previously defined with a nucleic acid selected from the group consisting of the nucleotide sequences 1-208, 1307-1350, 1703-1865, 2107-2180, 2843-3333, 3871-3882, 4222-4276, and 5017-5579 of SEQ ID No. 1 or a variant thereof or a sequence complementary thereto.

A nucleic probe or primer according to the invention comprises at least 8 consecutive nucleotides of a polynucleotide of SEQ ID Nos 1 or 4 or the sequences complementary thereto, preferably from 8 to 200 consecutive nucleotides, more particularly from 10, 15, 20 or 30 to 100 consecutive nucleotides, more preferably from 10 to 50 nucleotides, and most preferably from 40 to 50 consecutive nucleotides of a polynucleotide of SEQ ID Nos 1 or 4 or the sequences complementary thereto.

In a first preferred embodiment, the probe or primer is suspended in a suitable buffer for performing a hybridization or an amplification reaction.

In a second embodiment, the oligonucleotide probe, which may be immobilized on a support, is capable of hybridizing with a *RBP-7* gene, preferably with a region of the *RBP-7* gene which comprises a biallelic marker of the present invention. The techniques for immobilizing a nucleotide primer or probe on a solid support are well-known to the skilled artisan and include, but are not limited to, the immobilization techniques described in the present application.

In a third embodiment, the primer is complementary to any nucleotide sequence of the *RBP-7* gene and can be used to amplify a region of the *RBP-7* gene contained in the nucleic acid sample to be tested which includes a polymorphic base of at least one biallelic marker. Preferably, the amplified region includes a polymorphic base of at least one biallelic marker selected from the group consisting of SEQ ID Nos 30-71 or the sequences complementary

thereto. In some embodiments, the primer comprises one of the sequences of SEQ ID Nos 72-101 and 102-136.

When using a polynucleotide probe or primer in a detection method of the invention, the DNA or RNA contained in the sample to be assayed may be subjected to a first extraction step well known to the one skilled in the art, in order to make the DNA or RNA material contained in the initial sample available to a hybridization reaction, prior to the hybridization step itself.

The nucleic acid probes and primers of the invention are also used to detect and/or amplify a portion of the *RBP-7* gene within which a polymorphism or a mutation causes a change either in the expression level of the *RBP-7* gene or a change in the amino acid sequence of the *RBP-7* gene translation product.

The invention further concerns detection or amplification kits containing a pair of oligonucleotide primers or an oligonucleotide probe according to the invention. The kits of the present invention can also comprise optional elements including appropriate amplification reagents such as DNA polymerases when the kit comprises primers, or reagents useful in hybridization between a labeled hybridization probe and a *RBP-7* gene containing at least one biallelic marker. In one embodiment, the biallelic marker comprises one of the sequences of SEQ ID Nos 30-71 or the sequences complementary thereto.

In one embodiment the invention encompasses isolated, purified, and recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of any one of SEQ ID Nos 1 and 4 and the complement thereof, wherein said span includes a biallelic marker of RBP-7 in said sequence; optionally, wherein said biallelic marker of *RBP-7* is selected from the group consisting of A1 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said contiguous span is 18 to 47 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide; optionally, wherein said polynucleotide consists of or comprises said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide; optionally, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide; and optionally, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide. In a preferred embodiment, said probes comprises, consists of, or consists essentially of a sequence selected from the sequences SEQ ID Nos 30-71 and the complementary sequences thereto.

In another embodiment the invention encompasses isolated, purified and recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID Nos 1 and 4 or the complements thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of a biallelic marker of RBP-7 in said sequence; optionally, wherein said biallelic marker of RBP-7 is selected from the group consisting of A1 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said biallelic marker of RBP-7 in said sequence; and optionally, wherein said polynucleotide comprises, consists of, or consists essentially of a sequence selected from the sequences SEQ ID Nos 102-136.

In a further embodiment, the invention encompasses isolated, purified, or recombinant polynucleotides comprising, consisting of, or consisting essentially of a sequence selected from the sequences SEQ ID Nos 72-101.

In an additional embodiment, the invention encompasses polynucleotides for use in hybridization assays, sequencing assays, and enzyme-based mismatch detection assays for determining the identity of the nucleotide at a biallelic marker of RBP-7 in SEQ ID Nos 1 and 4, or the complements thereof, as well as polynucleotides for use in amplifying segments of nucleotides comprising a biallelic marker of RBP-7 in SEQ ID Nos 1 and 4, or the complements thereof; optionally, wherein said biallelic marker of RBP-7 is selected from the group consisting of A1 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith.

The formation of stable hybrids depends on the melting temperature (Tm) of the DNA. The Tm depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in the probes and primers of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %.

The length of these probes and probes can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 15 to 30 nucleotides. Shorter probes and primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes and primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art.

The primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang et al. (1979), the phosphodiester method of Brown et al. (1979), the diethylphosphoramidite method of Beaucage et al. (1981) and the solid support method described in EP 0 707 592, the disclosures of which are incorporated herein by reference in their entireties.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances (^{32}P, ^{35}S, ^{3}H, ^{125}I), fluorescent dyes (5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in the French Patent No. FR-7810975 or by Urdea et al (1988) or Sanchez-Pescador et al (1988). Advantageously, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al. in 1991 or in the European Patent No. EP-0225,807, the disclosure of which is incorporated herein by reference in its entirety (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect

PCR amplification products. They may also be used to detect mismatches in the *RBP-7* gene or mRNA using other techniques.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the inventions to a single solid support. In addition, polynucleotides other than those of the invention may attached to the same solid support as one or more polynucleotides of the invention.

Consequently, the invention also deals with a method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1, 4, a fragment or a variant thereof or the complementary sequence thereto in a sample, said method comprising the following steps of :

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes as described above and the sample to be assayed.

b) detecting the hybrid complex formed between the probe and a nucleic acid in the sample.

In a first preferred embodiment of this detection method, said nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule.

In a second preferred embodiment of said method, said nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate.

The invention further concerns a kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1, 4, a fragment or a variant thereof or the complementary sequence thereto in a sample, said kit comprising :

a) a nucleic acid probe or a plurality of nucleic acid probes as described above;

b) optionally, the reagents necessary for performing the hybridization reaction.

In a first preferred embodiment of the detection kit, the nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule.

In a second preferred embodiment of the detection kit, the nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate.

Oligonucleotide Arrays

A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in the *RBP-7* gene and may also be used for detecting mutations in the coding or in the non-coding sequences of the *RBP-7* gene.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT

publications WO 90/15070 and 92/10092, the disclosures of which are incorporated herein by reference in their entireties. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., Science, 251:767-777, 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, the disclosures of which are incorporated herein by reference in their entireties, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256, the disclosures of which are incorporated herein by reference in their entireties.

In another embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in the *RBP-7* gene and in its regulatory region. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion of substitution of one or several nucleotides). By known mutations is meant mutations on the *RBP-7* gene that have been identified according, for example to the technique used by Huang et al. (1996) or Samson et al. (1996).

Another technique that is used to detect mutations in the *RBP-7* gene is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of the *RBP-7* genomic DNA or cDNA. Thus, an array comprising, consisting essentially of, or consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild gene sequence, measure its amount, and detect differences between the target sequence and the reference wild gene sequence of the *RBP-7* gene. One such design, termed 4L tiled array, uses a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible

mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a "footprint" for the probes flanking a mutation position. This technique was described by Chee et al. in 1996, which is herein incorporated by reference.

Consequently, the invention concerns an array of nucleic acid comprising at least one polynucleotide described above as probes and primers. Preferably, the invention concerns an array of nucleic acid comprising at least two polynucleotides described above as probes and primers.

AMPLIFICATION OF THE *RBP-7* GENE

1. DNA Extraction

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples which can be tested by the methods of the present invention described herein and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the context of the present invention is from peripheral venous blood of each donor.

The techniques of DNA extraction are well-known to the skilled technician. Such techniques are described notably by Lin et al. (1998) and by Mackey et al. (1998).

2. DNA Amplification

DNA amplification techniques are well-known to those skilled in the art. Amplification techniques that can be used in the context of the present invention include, but are not limited to, the ligase chain reaction (LCR) described in EP-A- 320 308, WO 9320227 and EP-A-439 182, the disclosures of which are incorporated herein by reference, the polymerase chain reaction (PCR, RT-PCR) and techniques such as the nucleic acid sequence based amplification (NASBA) described in Guatelli JC, et al. (1990) and in Compton J. (1991), Q-beta amplification as described in European Patent Application No. 4544610, strand displacement amplification as described in Walker et al. (1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461, the disclosure of which is incorporated herein by reference.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes, which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall et al. (1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see White (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has

further been described in several patents including US Patents 4,683,195, 4,683,202 and 4,965,188. Each of these publications is incorporated by reference.

One of the aspects of the present invention is a method for the amplification of the human *RBP-7* gene, particularly of the genomic sequences of SEQ ID No. 1 or of the cDNA sequence of SEQ ID No. 4, or a fragment or a variant thereof in a test sample, preferably using the PCR technology. The method comprises the steps of contacting a test sample suspected of containing the target *RBP-7* encoding sequence or portion thereof with amplification reaction reagents comprising a pair of amplification primers, and eventually in some instances a detection probe that can hybridize with an internal region of amplicon sequences to confirm that the desired amplification reaction has taken place.

Thus, the present invention also relates to a method for the amplification of a human *RBP-7* gene sequence, particularly of a portion of the genomic sequences of SEQ ID No. 1 or of the cDNA sequence of SEQ ID No. 4, or a variant thereof in a test sample, said method comprising the steps of :

a) contacting a test sample suspected of containing the targeted *RBP-7* gene sequence comprised in a nucleotide sequence selected from a group consisting of SEQ ID Nos 1 and 4, or fragments or variants thereof with amplification reaction reagents comprising a pair of amplification primers as described above and located on either side of the polynucleotide region to be amplified, and

b) optionally detecting the amplification products.

In a preferred embodiment of the above amplification method, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region.

The primers are more particularly characterized in that they have sufficient complementarity with any sequence of a strand of the genomic sequence close to the region to be amplified, for example with a non-coding sequence adjacent to exons to amplify.

In a particular embodiment of the invention, the primers are selected form the group consisting of the nucleotide sequences detailed in Table C below.

TABLE C

Forward Primer Name	Position range of amplification primer in SEQ ID No. 1	Reverse Primer Name	Complementary position range of amplification primer in SEQ ID No. 1
P1	313-330	P26	732-751
P2	1282-1299	P27	1682-1699
P3	67531-67549	P28	67810-67830
P4	70927-70945	P29	71257-71276
P5	71613-71631	P30	72043-72060
P6	75390-75409	P31	75795-75814
P7	77544-77563	P32	77926-77943
P8	81708-81726	P33	82108-82127
P9	105046-105065	P34	105326-105345
P10	104751-104770	P35	105297-105316
P11	107691-107710	P36	108091-108110
P12	114296-114315	P37	114698-114716
P13	114327-114345	P38	114735-114753
P14	132101-132118	P39	132504-132521
P15	145522-145541	P40	145923-145942
P16	145866-145884	P41	146266-146285
P17	145956-145976	P42	146399-146418
P18	146529-146547	P43	146955-146972
P19	152763-152780	P44	153164-153182
P20	155404-155422	P45	155706-155726
P21	160043-160060	P46	160445-160462
P22	160361-160378	P47	160770-160788
P23	160742-160759	P48	161147-161165
P24	161127-161144	P49	161530-161547
P25	161217-161235	P50	161617-161636

The invention also concerns a kit for the amplification of a human *RBP-7* gene sequence, particularly of a portion of the genomic sequences of SEQ ID No. 1 or of the cDNA sequence of SEQ ID No. 4, or a variant thereof in a test sample, wherein said kit comprises :

a) A pair of oligonucleotide primers located on either side of the *RBP-7* region to be amplified;

b) Optionally, the reagents necessary for performing the amplification reaction.

In a preferred embodiment of the amplification kit described above, the primers are selected from the group consisting of the nucleotide sequences of SEQ ID Nos 72-101 and P1-P50.

In another embodiment of the above amplification kit, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region.

BIALLELIC MARKERS OF *RBP-7*

The inventors have discovered nucleotide polymorphisms located within the genomic DNA containing the *RBP-7* gene, and among them "Single Nucleotide Polymorphisms" or SNPs that are also termed biallelic markers.

The invention also relates to a nucleotide sequence, preferably a purified and/or isolated polynucleotide comprising a sequence defining a biallelic marker located in the sequence of a *RBP-7* gene, a fragment or variant thereof or a sequence complementary thereto. The sequences defining a biallelic marker may be of any length consistent with their intended use, provided that they contain a polymorphic base from a biallelic marker. Preferably, the sequences defining a biallelic marker include the polymorphic base of one of SEQ ID Nos 30-71 or the sequence complementary thereto. In some embodiments the sequences defining a biallelic marker comprise one of the sequences selected from the group consisting of SEQ ID Nos 30-71 or the sequences complementary thereto.

In a preferred embodiment, the invention relates to a set of purified and/or isolated nucleotide sequences, each sequence comprising a sequence defining a biallelic marker located in the sequence of a *RBP-7* gene, wherein the set is characterized in that between about 30 and 100%, preferably between about 40 and 60%, more preferably between 50 and 60%, of the sequences defining a biallelic marker are selected from the group consisting of SEQ ID Nos 30-71, the sequences complementary thereto, or a fragment or variant thereof.

The invention further concerns a nucleic acid encoding a RBP-7 protein, wherein said nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID Nos 30-71 or the sequences complementary thereto.

The invention also relates to nucleotide sequence selected from the group consisting of SEQ ID Nos 30-71, the sequences complementary thereto, or a fragment or a variant thereof.

A) Identification Of Biallelic Markers

There are two preferred methods through which the biallelic markers of the present invention can be generated.

In a first method, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms. One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained therewith usually shows a sufficient degree of informativeness for conducting association studies.

In a second method for generating biallelic markers, the DNA samples are not pooled and are therefore amplified and sequenced individually. The resulting nucleotide sequences obtained are then also analyzed to identify significant polymorphisms.

It will readily be appreciated that when this second method is used, a substantially higher number of DNA amplification reactions and sequencing reactions must be carried out. Moreover, a biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele may be less than about 10%. It will further be appreciated that including such less informative biallelic markers in association studies to identify potential genetic associations with a trait may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes.

The following is a description of the various parameters of a preferred method used by the inventors to generate the markers of the present invention.

1- DNA Extraction

The genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background.

The term "individual" as used herein refers to vertebrates, particularly members of the mammalian species and includes but is not limited to domestic animals, sports animals, laboratory animals, primates and humans. Preferably, the individual is a human.

The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, preferably from about 50 to about 200 individuals. It is usually preferred to collect DNA samples from at least about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples which can be tested by the methods of the present invention described herein and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic

DNA used in the context of the present invention is from peripheral venous blood of each donor.

The techniques of DNA extraction are well-known to the skilled technician. Details of a preferred embodiment are provided in Example 2.

Once genomic DNA from every individual in the given population has been extracted, it is preferred that a fraction of each DNA sample is separated, after which a pool of DNA is constituted by assembling equivalent amounts of the separated fractions into a single one. However, the person skilled in the art can choose to amplify the pooled or unpooled sequences

2- DNA Amplification

The identification of biallelic markers in a sample of genomic DNA may be facilitated through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well known to those skilled in the art. Various methods to amplify DNA fragments carrying biallelic markers are further described hereinbefore in "Amplification of the *RBP-7* gene". The PCR technology is the preferred amplification technique used to identify new biallelic markers. A typical example of a PCR reaction suitable for the purposes of the present invention is provided in Example 3.

In this context, one of the groups of oligonucleotides according to the present invention is a group of primers useful for the amplification of a genomic sequence encoding *RBP-7*. The primers pairs are characterized in that they have sufficient complementarity with any sequence of a strand of the *RBP-7* gene to be amplified, preferably with a sequence of introns adjacent to exons to amplify, with regions of the 3' and 5' ends of the *RBP-7* gene, with splice sites or with 5' UTRs or 3' UTRs to hybridize therewith.

These primers focus on exons and splice sites of the *RBP-7* gene since an identified biallelic marker as described below presents a higher probability to be an eventual causal mutation if it is located in these functional regions of the gene.

15 pairs of primers were designed with the aim of amplifying each of the 24 exons of the *RBP-7* gene (Table 1). To these primers can be added, at either end thereof, a further polynucleotide useful for sequencing such as described in Example 3. Preferred primers include those having the nucleotide sequences disclosed in Example 3. Some of the primers according to the invention allow the amplification of the majority of the *RBP-7* Exons shown in Figure 2.

The primers described above are individually useful as oligonucleotide probes in order to detect the corresponding *RBP-7* nucleotide sequence in a sample, and more preferably to detect the presence of a *RBP-7* DNA or RNA molecule in a sample suspected to contain it.

3- Sequencing Of Amplified Genomic DNA And Identification Of Polymorphisms

The amplification products generated as described above with the primers of the invention are then sequenced using methods known and available to the skilled technician. Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol.

Following gel image analysis and DNA sequence extraction, sequence data are automatically processed with adequate software to assess sequence quality

The sequence data obtained as described above are transferred to a database, where quality control and validation steps are performed. A base-caller, working using a Unix system automatically flags suspect peaks, taking into account the shape of the peaks, the inter-peak resolution, and the noise level. The base-caller also performs an automatic trimming. Any stretch of 25 or fewer bases having more than 4 suspect peaks is usually considered unreliable and is discarded.

After this first sequence quality analysis, polymorphism analysis software is used to detect the presence of biallelic sites among individual or pooled amplified fragment sequences. The polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern. These peaks, which present two distinct colors, correspond to two different nucleotides at the same position on the sequence. In order for peaks to be considered significant, peak height has to satisfy conditions of ratio between the peaks and conditions of ratio between a given peak and the surrounding peaks of the same color.

However, since the presence of two peaks can be an artifact due to background noise, two controls are utilized to exclude these artifacts :

- the two DNA strands are sequenced and a comparison between the peaks is carried out. The polymorphism has to be detected on both strands for validation.

- all the sequencing electrophoresis patterns of the same amplification product provided from distinct pools and/or individuals are compared. The homogeneity and the ratio of homozygous and heterozygous peak height are controlled through these distinct DNAs.

The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is about 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies. However, more than 90 % of the biallelic polymorphisms detected by the pooling method have a frequency for the minor allele higher than 0.25. Therefore, the biallelic markers selected by this method have a frequency of at least 0.1 for the minor allele and less than 0.9 for the major allele, preferably at least 0.2 for the minor allele and less than 0.8 for the major allele, more preferably at least 0.3 for the minor allele and less than

0.7 for the major allele, thus a heterozygosity rate higher than 0.18, preferably higher than 0.32, more preferably higher than 0.42.

In a particular embodiment of the invention, the test samples are a pool of 100 individuals and 50 individual samples. This is the methodology used in the preferred embodiment of the present invention, in which 21 biallelic markers have been identified in a genomic region containing the *RBP-7* gene. Their location on the genomic *RBP-7* DNA is shown in Figure 2 and their particular sequences are disclosed in example 4. The 24 exons and the intronic sequences surrounding the exons were analyzed. Among the 21 biallelic markers identified within the *RBP-7* gene, 6 biallelic markers are located within 4 different exons, and 15 biallelic markers are located within the different intronic regions. The biallelic markers 5-130-257, 5-143-84 and 5-143-101 respectively change asparagine into glycine, glycine into glutamic acid and leucine into methionine in the RBP-7 protein. The amino acid changes caused by the 5-143-84 biallelic marker may be important for the RBP-7 biological activity, since a neutral amino acid is replaced by a positively charged amino acid in a RBP-7 region likely to contain a domain involved in a non-covalent interaction with the retinoblastoma protein or also a pRb related protein such as p107 or p130.

4- Validation Of The Biallelic Markers Of The Present Invention

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals, so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a sequence. All of the genotyping, haplotyping, and association study methods of the invention may optionally be performed solely with validated biallelic markers.

5- Evaluation Of The Frequency Of The Biallelic Markers Of The Present Invention

The validated biallelic markers are further evaluated for their usefulness as genetic markers by determining the frequency of the least common allele at the biallelic marker site. The higher the frequency of the less common allele the greater the usefulness of the biallelic marker in association and interaction studies. The determination of the least common allele is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. This determination of frequency by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group must be large enough to be representative of the population as a whole. Preferably the group contains at least 20 individuals, more preferably the group contains at least 50 individuals, most preferably the group contains at least 100 individuals. Of course the larger the group the greater the accuracy of the frequency determination because of reduced sampling error. A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker." All of the genotyping, haplotyping, and association interaction study methods of the invention may optionally be performed solely with high quality biallelic markers.

B- Genotyping An Individual For Biallelic Markers

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which may be performed *in vitro*. Such methods of genotyping comprise determining the identity of a nucleotide at an *RBP-7* biallelic marker site by any method known in the art. These methods find use in genotyping case-control populations in association studies as well as individuals in the context of detection of alleles of biallelic markers which are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual may be classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed nucleic acid samples derived from a single individual or pooled DNA samples.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

1- Source Of DNA For Genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described above in " DNA extraction". While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

2- Amplification Of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention.

Amplification of DNA may be achieved by any method known in the art. Amplification techniques are described above under the headings "Amplification of the *RBP-7* gene".

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described below.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention.

In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. Preferred amplification primers are listed in Example 3. It will be appreciated that the primers listed are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention.

The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention amplified segments carrying biallelic markers can range in size

from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers may be any sequence which allow the specific amplification of any DNA fragment carrying the markers. Amplification primers may be labeled or immobilized on a solid support as described under the headings entitled " Oligonucleotide probes and primers".

<u>3- Methods Of Genotyping DNA Samples For Biallelic Markers</u>

<u>a- Sequencing assays</u>

The amplification products generated above with the primers of the invention can be sequenced using methods known and available to the skilled technician. Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. A sequence analysis can allow the identification of the base present at the polymorphic site.

<u>b- Microsequencing assays</u>

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883, the disclosure of which is incorporated herein by reference in its entirety. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in Example 5.

Different approaches can be used for the labeling and detection of ddNTPs. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (1997) and Chen et al. (1997). In this method amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein–labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without

separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time.

Microsequencing may be achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is conducted as a heterogenous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator regent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner oligonucleotides or templates may be attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, 1994) or linked to fluorescein (Livak and Hainer, 1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712, the disclosure of which is incorporated herein by reference in its entirety). As yet another alternative solid-phase microsequencing procedure, Nyren et al. (1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen et al. (1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an

oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described below.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. Preferred microsequencing primers include those being featured in Example 5. It will be appreciated that the microsequencing primers listed in Example 5 are merely exemplary and that, any primer having a 3' end immediately adjacent to the polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for any biallelic marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Example 5, or fragments comprising at least 8, at least 12, at least 15, or at least 20 consecutive nucleotides thereof and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a nucleotide at a biallelic marker site.

c- Mismatch Detection Assays Based On Polymerases And Ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by allele-specific amplification assays. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above.

Allele specific amplification primers

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. This can be accomplished by placing the polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3'end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Determining the precise location of the mismatch and the corresponding assay conditions are well within the ordinary skill in the art.

Ligation/Amplification Based Methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting single nucleotide polymorphisms and may be

advantageously combined with PCR as described by Nickerson et al. (1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other amplification methods which are particularly suited for the detection of single nucleotide polymorphism include LCR (ligase chain reaction), Gap LCR (GLCR) which are described above in "Amplification of the *RBP-7* gene". LCR uses two pairs of probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO 90/01069, the disclosure of which is incorporated herein by reference in its entirety. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit Analysis™ is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271, the disclosure of which is incorporated herein by reference in its entirety). This method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

d- Hybridization Assay Methods

A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. Although such hybridizations can be performed in solution, it is preferred to employ a solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes. Preferably, the hybrids can be bound to a solid phase reagent by virtue of a capture label and detected by virtue of a detection label. In cases where the detection label is directly detectable, the presence of the hybrids on the solid phase can be detected by causing the label to produce a detectable signal, if necessary, and detecting the signal. In cases where the label is not directly detectable, the captured hybrids can be contacted with a conjugate, which generally comprises a binding member attached to a directly detectable label. The conjugate becomes bound to the complexes and the conjugates presence on the complexes can be detected with the directly detectable label. Thus, the presence of the hybrids on the solid phase reagent can be determined.

The polynucleotides provided herein can be used to produce probes which can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes are characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. A particularly preferred probe is 25 nucleotides in length. Preferably the polymorphic site of the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes the polymorphic site of the biallelic marker is at the center of said polynucleotide.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in "Oligonucleotide probes and primers". Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Nos. 5,185,444; 5,034,506 and 5,142,047, the disclosures of which are incorporated herein by reference in their entireties. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is No. longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

The probes of the present invention are useful for a number of purposes. By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample.

High-Throughput parallel hybridizations in array format are specifically encompassed within "hybridization assays" and are described below.

e- Hybridization To Addressable Arrays Of Oligonucleotides

DNA chips result from the adaptation of computer chips to biology. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S.cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia et al., 1996; Shoemaker et al., 1996 ; Kozal et al., 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP785280, the disclosure of which is incorporated herein by reference in its entirety, describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of monomers, i.e. nucleotides. Tiling strategies are further described in PCT Application No. WO 95/11995, the disclosure of which is incorporated herein by reference in its entirety. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of biallelic markers. For example, a detection block may be tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the polymorphic site of the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in

PCT Application No. WO 92/10092 and WO 95/11995 and US Patent No. 5,424,186, the disclosures of which are incorporated herein by reference in their entireties.

Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences of fragments of about 15 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of the nucleic acids of the sequences set forth as SEQ ID Nos 30-75 and the sequences complementary thereto, or a fragment thereof at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, or 40 consecutive nucleotides comprising a biallelic marker of the present invention. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more of these polynucleotides of the invention. Solid supports and polynucleotides of the present invention attached to solid supports are further described in "Oligonucleotide primers and probes".

f- Integrated Microsequencing And Capillary Electrophoresis Chips

Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US Patent 5,589,136, the disclosure of which is incorporated herein by reference in its entirety, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts.

For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

ASSOCIATION STUDIES WITH THE BIALLELIC MARKERS OF THE *RBP-7* GENE

The identification of genes involved in suspected heterogeneous, polygenic and multifactorial traits such as cancer can be carried out through two main strategies currently used for genetic mapping: linkage analysis and association studies. Association studies examine the frequency of marker alleles in unrelated trait positive (T+) individuals compared with trait negative (T-) controls, and are generally employed in the detection of polygenic inheritance.

Association studies as a method of mapping genetic traits rely on the phenomenon of linkage disequilibrium, which is described below.

If two genetic loci lie on the same chromosome, then sets of alleles of these loci on the same chromosomal segment (called haplotypes) tend to be transmitted as a block from generation to generation. When not broken up by recombination, haplotypes can be tracked not only through pedigrees but also through populations. The resulting phenomenon at the population level is that the occurrence of pairs of specific alleles at different loci on the same chromosome is not random, and the deviation from random is called linkage disequilibrium (LD).

If a specific allele in a given gene is directly involved in causing a particular trait T, its frequency will be statistically increased in a T+ population when compared to the frequency in a T- population. As a consequence of the existence of LD, the frequency of all other alleles present in the haplotype carrying the trait-causing allele (TCA) will also be increased in T+ individuals compared to T- individuals. Therefore, association between the trait and any allele in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular allele's region. Linkage disequilibrium allows the relative frequencies in T+ and T- populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analyzed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles.

The general strategy to perform association studies using biallelic markers derived from a candidate region is to scan two groups of individuals (trait + and trait – control individuals which are characterized by a well defined phenotype as described below) in order to measure and statistically compare the allele frequencies of such biallelic markers in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analyzed biallelic markers, one can assume that : either the associated allele is directly responsible for causing the trait (associated allele is the TCA), or the associated allele is in LD with the TCA. If the evidence indicates that the associated allele within the candidate region is most probably not the TCA but is in LD with the real TCA, then the TCA, and by consequence the gene carrying the TCA, can be found by sequencing the vicinity of the associated marker.

It is another object of the present invention to provide a method for the identification and characterization of an association between alleles for one or several biallelic markers of the human *RBP-7* gene and a trait. The method comprises the steps of :

- genotyping a marker or a group of biallelic markers according to the invention in trait positive and trait negative individuals; and

- establishing a statistically significant association between one allele of at least one marker and the trait.

Preferably, the trait positive and trait negative individuals are selected from non-overlapping phenotypes, at opposite ends of the non-bimodal phenotype spectra of the trait under study. In some embodiments, the biallelic marker is one of the biallelic markers of the present invention.

In a preferred embodiment, the trait is a disease and preferably a cancer.

The present invention also provides a method for the identification and characterization of an association between a haplotype comprising alleles for several biallelic markers of the human *RBP-7* gene and a trait. The method comprises the steps of :

- genotyping a group of biallelic markers according to the invention in trait positive and trait negative individuals; and

- establishing a statistically significant association between a haplotype and the trait.

In some embodiments, the haplotype comprises two or more biallelic markers defined in SEQ ID Nos 30-71.

The step of testing for and detecting the presence of DNA comprising specific alleles of a biallelic marker or a group of biallelic markers of the present invention can be carried out as described further below.

VECTORS FOR THE EXPRESSION OF A REGULATORY OR A CODING POLYNUCLEOTIDE ACCORDING TO THE INVENTION

Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences, coding sequences and polynucleotide constructs, as well as any *RBP-7* primer or probe as defined above. More particularly, the recombinant vectors of the present invention can comprise any of the polynucleotides described in the "*RBP-7* Gene, Corresponding cDNAs And *RBP-7* Coding And Regulating Sequences" section, and the "Oligonucleotide Probes And Primers" section.

Any of the regulatory polynucleotides or the coding polynucleotides of the invention may be inserted into recombinant vectors for expression in a recombinant host cell or a recombinant host organism.

Thus, the present invention also encompasses a family of recombinant vectors that contains either a *RBP-7* regulatory polynucleotide or a *RBP-7* coding polynucleotide or both of them. Preferably, the present invention concerns recombinant vectors that contains either a *RBP-7* regulatory polynucleotide or a *RBP-7* coding polynucleotide comprising at least one of the biallelic markers of the invention, particularly those of SEQ ID Nos 30-71.

More particularly, the present invention also relates to expression vectors which include nucleic acids encoding a *RBP-7* protein under the control of either a *RBP-7* regulatory polynucleotide, or an exogenous regulatory sequence.

Another aspect of the present invention is a recombinant expression vector comprising a nucleic acid selected from the group consisting of SEQ ID Nos 1, 4, 5-28 or complementary sequences thereto or fragments or variants thereof.

Another preferred recombinant expression vector according to the invention comprises a nucleic acid comprising a combination of at least two polynucleotides selected from the group consisting of SEQ ID Nos 5-28 or the sequences complementary thereto, wherein the polynucleotides are arranged within the nucleic acid, from the 5' end to the 3'end of said nucleic acid, in the same order than in the SEQ ID No. 1.

Another aspect of the invention is a recombinant expression vector comprising a nucleic acid selected from the group consisting of SEQ ID No. 2 or 3 or the sequences complementary thereto or a biologically active fragment or variant thereof.

A further aspect of the invention is a recombinant expression vector comprising a purified or isolated nucleic acid comprising :

a) a nucleic acid comprising the nucleotide sequence SEQ ID No. 2, a fragment or variant thereof or a nucleotide sequence complementary thereto;

b) a polynucleotide encoding a protein or a polynucleotide of interest.

The invention also encompasses a recombinant expression vector containing a polynucleotide comprising, consisting essentially of, or consisting of :

a) a nucleic acid comprising a regulatory polynucleotide of SEQ ID No. 2, or the sequence complementary thereto , or a biologically active fragment or variant thereof; and

b) a polynucleotide encoding a polypeptide or a polynucleotide of interest.

c) Optionally, the expression vector may further comprise a nucleic acid comprising a regulatory polynucleotide of SEQ ID No. 3, or the sequence complementary thereto, or a biologically active fragment or variant thereof.

The vector containing the appropriate DNA sequence as described above, more preferably a *RBP-7* regulatory polynucleotide, a *RBP-7* coding polynucleotide or both of them, can be utilized to transform an appropriate host to allow the expression of the desired polypeptide or polynucleotide.

Vectors

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may comprise, consist

essentially of, or consist of a chromosomal, non-chromosomal and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of

(1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, and

(3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The selectable marker genes can be for example dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicine or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired polypeptide with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA).

A suitable vector for the expression of the RBP-7 protein above-defined or their peptide fragments is a baculovirus vector that can be propagated in insect cells and in insect cell lines. A specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector

(Pharmingen) that is used to transfect the SF9 cell line (ATCC No. CRL 1711) which is derived from *Spodoptera frugiperda*. Other baculovirus vectors are described in Chai et al. (1993), Vlasak et al. (1983) and Lenhardt et al. (1996).

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example SV40 origin, early promoter, enhancer, splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Large numbers of suitable vectors and promoters are known to those of skill in the art, and commercially available, such as bacterial vectors : pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); or eukaryotic vectors : pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); baculovirus transfer vector pVL1392/1393 (Pharmingen); pQE-30 (QIAexpress).

Promoters

The suitable promoter regions used in the expression vectors according to the present invention are choosen taking into account of the cell host in which the heterologous gene has to be expressed.

Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith et al., 1983.; O'Reilly et al., 1992), the lambda P_R promoter or also the trc promoter.

Preferred promoters for the expression of the heterologous gene in eukaryotic hosts are the early promoter of CMV, the Herpes simplex virus thymidine kinase promoter, the early or the late promoter from SV40, the LTR regions of certain retroviruses or also the mouse metallothionein I promoter.

Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors. Particularly named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, PL and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

The choice of a determined promoter, among the above-described promoters is well in the ability of one skill in the art, guided by his knowledge in the genetic engineering technical field, and by being also guided by the book of Sambrook et al. in 1989 or also by the procedures described by Fuller et al. in 1996.

Other Types Of Vectors

The *in vivo* expression of a RBP-7 polypeptide or a fragment or a variant thereof may be useful in order to study the physiological consequences of a deregulation of its *in vivo* synthesis on the physiology of the recipient recombinant host organism under study, more particularly on the cell differentiation and on an eventual abnormal proliferation of various kinds of cells, including T cells and epithelial cells.

Consequently, the present invention also relates to recombinant expression vectors mainly designed for the *in vivo* production of a therapeutic peptide fragment by the introduction of the genetic information in the organism of the patient to be treated. This genetic information may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue.

The method for delivering the corresponding protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect.

In a specific embodiment, the invention provides a composition for the *in vivo* production of a RBP-7 polypeptide containing a naked polynucleotide operatively coding for a RBP-7 polypeptide or a fragment or a variant thereof, in solution in a physiologically acceptable carrier and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

Advantageously, the composition described above is administered locally, near the site in which the expression of a RBP-7 polypeptide or a fragment or a variant thereof is sought.

The polynucleotide operatively coding for a RBP-7 polypeptide or a fragment or variant thereof may be a vector comprising the genomic DNA or the complementary DNA (cDNA) coding for the corresponding protein or its protein derivative and a promoter sequence allowing the expression of the genomic DNA or the complementary DNA in the desired eukaryotic cells, such as vertebrate cells, specifically mammalian cells.

The promoter contained in such a vector is selected among the group comprising :

- an internal or an endogenous promoter, such as the natural promoter associated with the structural gene coding for the desired RBP-7 polypeptide or the fragment or variant thereof; such a promoter may be completed by a regulatory element derived from the vertebrate host, in particular an activator element;

- a promoter derived from a cytoskeletal protein gene such as the desmin promoter (Bolmont et al., 1990; Zhenlin et al., 1989).

As a general feature, the promoter may be heterologous to the vertebrate host, but it is advantageously homologous to the vertebrate host.

By a promoter heterologous to the vertebrate host is intended a promoter that is not found naturally in the vertebrate host.

Compositions comprising a polynucleotide are described in the PCT Application No. WO 90/11092 and also in the PCT Application No. WO 95/11307 as well as in the articles of Tacson et al. (1996) and of Huygen et al. (1996), the disclosures of which are incorporated herein by reference in their entireties.

In another embodiment, the DNA to be introduced is complexed with DEAE-dextran (Pagano et al., 1967) or with nuclear proteins (Kaneda et al., 1989), with lipids (Felgner et al., 1987) or encapsulated within liposomes (Fraley et al., 1980).

In another embodiment, the polynucleotide encoding a RBP-7 polypeptide or a fragment or a variant thereof may be included in a transfection system comprising polypeptides that promote its penetration within the host cells as it is described in the PCT Application WO 95/10534, the disclosure of which is incorporated herein by reference in its entirety.

The vector according to the present invention may advantageously be administered in the form of a gel that facilitates their transfection into the cells. Such a gel composition may be a complex of poly-L-lysine and lactose, as described by Midoux (1993) or also poloxamer 407 as described by Pastore (1994). Said vector may also be suspended in a buffer solution or be associated with liposomes.

The amount of the vector to be injected to the desired host organism vary according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 μg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, said vector may be introduced in vitro in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired RBP-7 polypeptide or the desired fragment or variant thereof is implanted back into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

Suitable vectors for the *in vivo* expression of a RBP-7 polypeptide or a fragment or a variant thereof are described hereunder.

In one specific embodiment, the vector is derived from an adenovirus. Preferred adenovirus vectors according to the invention are those described by Feldman and Steg (1996) or Ohno et al. (1994). Another preferred recombinant adenovirus according to this specific embodiment of the present invention is the adenovirus described by Ohwada et al. (1996) or the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French Patent Application No. FR-93.05954, the disclosure of which is incorporated herein by reference in its entirety).

Among the adenoviruses of animal origin it can be cited the adenoviruses of canine (CAV2, strain Manhattan or A26/61[ATCC VR-800]), bovine, murine (Mav1, Beard et al., 1980) or simian (SAV). Other adenoviruses are described by Levrero et al. (1991), Graham et al. (1984), in the European Patent Application No. EP-185.573 or in the PCT Application No. WO 95/14785, the disclosures of which are incorporated herein by reference in their entireties.

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous polynucleotides *in vivo* , particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host. Suitable retroviruses used according to the present invention include those described in the PCT Application No. WO 93/25234, the PCT Application No. WO 94/06920, the PCT Application No. WO 94/ 24298, Roth et al. (1996), Roux et al. (1989), Julian et al. (1992) and Neda et al. (1991), the disclosures of which are incorporated herein by reference in their entireties. Other preferred retrovirus include Murine Leukemia Viruses such as 4070A and 1504A (Hartley et al., 1976), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Gross (ATCC No. VR-590), Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190; PCT Application No. WO 94/24298), the disclosure of which is incorporated herein by reference in its entirety, and also Rous Sarcoma Viruses such as Bryan high titer (ATCC Nos VR-334, VR-657, VR-726, VR-659 and VR-728.

Yet another viral vector system that is contemplated by the invention comprises the adeno-associated virus (AAV). Adeno-associated virus is a naturally occuring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka et al., 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte et al., 1992; Samulski et al., 1989; McLaughlin et al., 1989). One

advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

Other compositions containing a vector of the invention comprise advantageously an oligonucleotide fragment of the nucleic sequence of *RBP-7* as an antisense tool that inhibits the expression of the corresponding gene and is thus useful to inhibit the expression of the *RBP-7* gene in the tagged cells or organs. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al. (1995) or also in the PCT Application No. WO 95/24223, the disclosure of which is incorporated herein by reference in its entirety.

Vectors suitable for homologous recombination

Other suitable vectors, particularly for the expression of genes in mammalian cells, may be selected from the group of vectors consisting of P1 bacteriophages, and bacterial artificial chromosomes (BACs). These types of vectors may contain large inserts ranging from about 80-90 kb (P1 bacteriophage) to about 300 kb (BACs).

P1 bacteriophage.

The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg (1992, 1994). Recombinant P1 clones comprising *RBP-7* nucleotide sequences may be designed for inserting large polynucleotides of more than 40 kb (Linton et al., 1993). To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick et al. (1994). Briefly, *E. coli* (preferably strain NS3529) harboring the P1 plasmid are grown overnight in a suitable broth medium containing 25 μg/ml of kanamycin. The P1 DNA is prepared from the *E. coli* by alkaline lysis using the Qiagen Plasmid Maxi kit (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

When the goal is to express a P1 clone comprising *RBP-7* nucleotide sequences in a transgenic animal, typically in transgenic mice, it is desirable to remove vector sequences from the P1 DNA fragment, for example by cleaving the P1 DNA at rare-cutting sites within the P1 polylinker (*Sfi*I, *Not*I or *Sal*I). The P1 insert is then purified from vector sequences on a pulsed-field agarose gel, using methods similar using methods similar to those originally reported for the isolation of DNA from YACs (Schedl et al., 1993a; Peterson et al., 1993). At this stage, the resulting purified insert DNA can be concentrated, if necessary, on a Millipore Ultrafree-MC Filter Unit (Millipore, Bedford, MA, USA – 30,000 molecular weight limit) and then dialyzed

against microinjection buffer (10 mM Tris-HCl, pH 7.4; 250 μM EDTA) containing 100 mM NaCl, 30 μM spermine, 70 μM spermidine on a microdyalisis membrane (type VS, 0.025 μM from Millipore). The intactness of the purified P1 DNA insert is assessed by electrophoresis on 1% agarose (Sea Kem GTG; FMC Bio-products) pulse-field gel and staining with ethidium bromide.

Bacterial Artificial Chromosomes (BACs)

The bacterial artificial chromosome (BAC) cloning system (Shizuya et al., 1992) has been developed to stably maintain large fragments of genomic DNA (100-300 kb) in *E. coli.* A preferred BAC vector is the pBeloBAC11 vector that has been described by Kim et al. (1996) BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into either the *Bam* HI or *Hind*III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites that can be used to generate end probes by either RNA transcription or PCR methods. After the construction of a BAC library in *E. coli*, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The cloning site is flanked by two *Not* I sites, permitting cloned segments to be excised from the vector by *Not* I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector may be linearized by treatment of the BAC vector with the commercially available enzyme lambda terminase that leads to the cleavage at the unique *cos*N site, but this cleavage method results in a full length BAC clone containing both the insert DNA and the BAC sequences.

Specific DNA construct vector for homologous recombination

The term "DNA construct" is understood to mean a linear or circular purified or isolated polynucleotide that has been artificially designed and which comprises at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their natural environment.

DNA CONSTRUCT THAT ENABLES DIRECTING TEMPORAL AND SPATIAL GENE EXPRESSION IN RECOMBINANT CELL HOSTS AND IN TRANSGENIC ANIMALS

In order to study the physiological and phenotype consequences of a lack of synthesis of the RBP-7 protein, both at the cell level and at the multi cellular organism level, in particular as regards to disorders related to abnormal cell proliferation, notably cancers, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of the *RBP-7* genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as

regards to the *RBP-7* nucleotide sequence of SEQ ID Nos 1 or 4, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon , an intron or a regulatory sequence, but preferably in the 5'-regulatory sequence or in an exon of the *RBP-7* genomic sequence or within the *RBP-7* cDNA of SEQ ID No. 4.

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E. coli* transposon Tn110 for controlling the *RBP-7* gene expression, such as described by Gossen et al. (1992, 1995) and Furth et al. (1994). Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tet*op) that are fused to either a minimal promoter or a 5'-regulatory sequence of the *RBP-7* gene, said minimal promoter or said *RBP-7* regulatory sequence being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a RBP-7 polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention will comprise both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor.

In the specific embodiment wherein the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

DNA Constructs Allowing Homologous Recombination : Replacement Vectors

A second preferred DNA construct will comprise, from 5'-end to 3'-end : (a) a first nucleotide sequence that is comprised in the *RBP-7* genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence that is comprised in the *RBP-7* genomic sequence, and is located on the genome downstream the first *RBP-7* nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (b). Preferably, the negative selection marker is the thymidine kinase (*tk*) gene (Thomas et al., 1986), the hygromycine beta gene (Te Riele et al., 1990), the *hprt* gene (Van der Lugt et al., 1991; Reid et al., 1990) or the Diphteria toxin A fragment (*Dt-A*) gene (Nada et al., 1993; Yagi et al. 1990). Preferably, the positive selection marker is located within a *RBP-7* exon sequence so as to interrupt the sequence encoding a *RBP-7* protein.

These replacement vectors are described for example by Thomas et al. (1986; 1987), Mansour et al. (1988) and Koller et al. (1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a *RBP-7* regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) is ranging from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

DNA Constructs Allowing Homologous Recombination : Cre-Loxp System

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *lox*P site. The *lox*P site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess et al., 1986). The recombination by the Cre enzyme between two *lox*P sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*lox*P system used in combination with a homologous recombination technique has been first described by Gu et al. (1993, 1994). Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *lox*P sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki et al. (1995), or by lipofection of the enzyme into the cells, such as described by Baubonis et al. (1993); (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being introduced in the recombinant cell host, such as described by Gu et al. (1993) and Sauer et al. (1988); (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu et al. (1994).

In the specific embodiment wherein the vector containing the sequence to be inserted in the *RBP-7* gene by homologous recombination is constructed in such a way that selectable markers are flanked by *lox*P sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the *RBP-7* sequences of interest that

have been inserted by an homologous recombination event. Again, two selectable markers are needed : a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-*lox*P system are described by Zou et al. (1994).

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end : (a) a first nucleotide sequence that is comprised in the *RBP-7* genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *lox*P site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised in the *RBP-7* genomic sequence, and is located on the genome downstream of the first *RBP-7* nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *lox*P site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *lox*P sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two *lox*P sites, is performed at a desired time, due to the presence within the genome of the recombinant cell host of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu et al. (1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result of the breeding of two transgenic animals, the first transgenic animal bearing the *RBP-7*-derived sequence of interest containing the *lox*P sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu et al. (1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton and Graham (1995) and Kanegae et al. (1995).

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a *RBP-7* genomic sequence or a *RBP-7* cDNA sequence, and most preferably an altered copy of a *RBP-7* genomic or cDNA sequence, within a

predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination).

Nuclear Antisense DNA Constructs

Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu et al. (1994). In a preferred embodiment, these *RBP-7* antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation , such as described by Eckner et al. (1991).

CELL HOSTS

Another aspect of the invention is a host cell that has been transformed or transfected with one of the polynucleotides described herein, and in particular a polynucleotide either comprising a *RBP-7* regulatory polynucleotide or the coding sequence of the *RBP-7* polypeptide selected from the group consisting of SEQ ID Nos 1 and 4 or a fragment or a variant thereof. Also included are host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above. More particularly, the cell hosts of the present invention can comprise any of the polynucleotides described in the "*RBP-7* Gene, Corresponding cDNAs And *RBP-7* Coding And Regulating Sequences" section, and the "Oligonucleotide Probes And Primers" section.

A further recombinant cell host according to the invention comprises a polynucleotide containing a biallelic marker selected from the group consisting of A1 to A21, and the complements thereof.

An additional recombinant cell host according to the invention comprises any of the vectors described herein, more particularly any of the vectors described in the " Vectors For The Expression Of A Regulatory Or A Coding Polynucleotide According To The Invention " section.

All the above-described vectors are useful to transform or transfect cell hosts in order to express a polynucleotide coding for a RBP-7 polypeptide or their peptide fragments or variants, or a polynucleotide of interest derived from the *RBP-7* gene.

Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, as well as various species within the genera of *Streptomyces* or *Mycobacterium*. Suitable eukaryotic hosts comprise yeast, insect cells, such as Drosophila and Sf9. Various mammalian cell hosts

can also be employed to express recombinant protein. Examples of mammalian cell hosts include the COS-7 lines of monkey kidney fibroblasts (Guzman, 1981), and other cell lines capable of expressing a compatible vector, for example the C127, 3T3, CHO, HeLa and BHK cell lines. The selection of an host is within the scope of the one skilled in the art.

A cell host according to the present invention is characterized in that its genome or genetic background (including chromosome, plasmids) is modified by the heterologous nucleic acid coding for a RBP-7 polypeptide or a peptide fragment or variant, or by a polynucleotide of interest derived from the *RBP-7* gene.

Preferred cell hosts used as recipients for the expression vectors of the invention are the followings :

a) Prokaryotic cells : *Escherichia coli* strains (I.E. DH5-□ strain) or *Bacillus subtilis*.

b) Eukaryotic cell hosts : HeLa cells (ATCC No. CCL2; No. CCL2.1; No. CCL2.2), Cv 1 cells (ATCC No. CCL70), COS cells (ATCC No. CRL1650; No. CRL1651), Sf-9 cells (ATCC No. CRL1711), mammal ES stem cells.

Preferably, the mammal ES stem cells include human (Thomson et al., 1998), mice, rats and rabbits ES stem cells and are preferably used in a process for producing transgenic animals, such as those described below.

The *RBP-7* gene expression in human cells may be rendered defective, or alternatively it may be proceeded with the insertion of a *RBP-7* genomic or cDNA sequence with the replacement of the *RBP-7* gene counterpart in the genome of an animal cell by a *RBP-7* polynucleotide according to the invention. These genetic alterations may be generated by homologous recombination events using specific DNA constructs that have been previously described.

One kind of cell hosts that may be used are mammal zygotes, such as murine zygotes. For example, murine zygotes may undergo microinjection with a purified DNA molecule of interest, for example a purified DNA molecule that has previously been adjusted to a concentration range from 1 ng/ml –for BAC inserts- 3 ng/μl –for P1 bacteriophage inserts- in 10 mM Tris-HCl, pH 7.4, 250 μM EDTA containing 100 mM NaCl, 30 μM spermine, and70 μM spermidine. When the DNA to be microinjected has a large size, polyamines and high salt concentrations can be used in order to avoid mechanical breakage of this DNA, as described by Schedl et al (1993b).

Anyone of the polynucleotides of the invention, including the DNA constructs described herein, may be introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommited cells of the inner cell mass of pre-implantation blastocysts. Peferred ES cell lines are the following : ES-E14TG2a (ATCC

No. CRL-1821), ES-D3 (ATCC No. CRL1934 and No. CRL-11632), YS001 (ATCC No. CRL-11776), 36.5 (ATCC No. CRL-11116). To maintain ES cells in an uncommitted state, they are cultured in the presence of growth inhibited feeder cells which provide the appropriate signals to preserve this embryonic phenotype and serve as a matrix for ES cell adherence. Preferred feeder cells are primary embryonic fibroblasts that are established from tissue of day 13- day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo et al. (1993) and are inhibited in growth by irradiation, such as described by Robertson (1987), or by the presence of an inhibitory concentration of LIF, such as described by Pease and Williams (1990).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

TRANSGENIC ANIMALS

The terms "transgenic animals" or "host animals" are used herein designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a *RBP-7* coding sequence, a *RBP-7* regulatory polynucleotide or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

Preferred transgenic animals according to the invention contains in their somatic cells and/or in their germ line cells any one of the polynucleotides, the recombinant vectors and the cell hosts described in the present invention. More particularly, the transgenic animals of the present invention can comprise any of the polynucleotides described in the "*RBP-7* Gene,

Corresponding cDNAs And *RBP-7* Coding And Regulating Sequences" section, the "Oligonucleotide Probes And Primers" section, the " Vectors For The Expression Of A Regulatory Or A Coding Polynucleotide According To The Invention " section and the "Cell Hosts" section.

The transgenic animals of the invention thus contain specific sequences of exogenous genetic material such as the nucleotide sequences described above in detail.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to cell differentiation, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native RBP-7 protein, or alternatively a mutant RBP-7 protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the *RBP-7* gene, leading to good yields in the synthesis of this protein of interest, and eventually a tissue specific expression of this protein of interest.

The design of the transgenic animals of the invention may be made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to Sandou et al. (1994) and also to US Patents Nos 4,873,191, issued Oct.10, 1989, 5,464,764 issued Nov 7, 1995 and 5,789,215, issued Aug 4, 1998, these documents being herein incorporated by reference to disclose methods producing transgenic mice.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a *RBP-7* coding sequence, a *RBP-7* regulatory polynucleotide or a DNA sequence encoding a *RBP-7* antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas et al. (1987). The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour et al. (1988).

Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley (1987). The blastocysts are then inserted into a female host animal and allowed to grow to term.

Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood et al. (1993) or by Nagy et al. (1993), the ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line.

The offsprings of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention

A further aspect of the invention is recombinant cell hosts obtained from a transgenic animal described herein.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989) and Shay et al. (1991).

RBP-7 POLYPEPTIDES

It is now easy to produce proteins in high amounts by genetic engineering techniques through expression vectors such as plasmids, phages or phagemids. The polynucleotide that code for one the polypeptides of the present invention is inserted in an appropriate expression vector in order to produce *in vitro* the polypeptide of interest.

Thus, the present invention also concerns a method for producing one of the polypeptides described herein, and especially a polypeptide of SEQ ID No. 29 or a fragment or a variant thereof, wherein said method comprises the steps of :

a) Optionally amplifying the nucleic acid coding for a RBP-7 polypeptide, or a fragment or a variant thereof, using a pair of primers according to the invention (by PCR, SDA, TAS, 3SR NASBA, TMA etc.).

b) Inserting the resulting amplified nucleic acid in an appropriate vector;

c) culturing, in an appropriate culture medium, a cell host previously transformed or transfected with the recombinant vector of step b);

d) harvesting the culture medium thus conditioned or lyse the cell host, for example by sonication or by an osmotic shock;

e) separating or purifying, from the said culture medium, or from the pellet of the resultant host cell lysate the thus produced polypeptide of interest.

f) Optionally characterizing the produced polypeptide of interest.

The polypeptides according to the invention may be characterized by binding onto an immunoaffinity chromatography column on which polyclonal or monoclonal antibodies directed to a polypeptide of SEQ ID No. 29, or a fragment or a variant thereof, have previously been immobilized.

Purification of the recombinant proteins or peptides according to the present invention may be carried out by passage onto a Nickel or Cupper affinity chromatography column. The Nickel chromatography column may contain the Ni-NTA resin (Porath et al., 1975).

The polypeptides or peptides thus obtained may be purified, for example by high performance liquid chromatography, such as reverse phase and/or cationic exchange HPLC, as described by Rougeot et al. (1994). The reason to prefer this kind of peptide or protein purification is the lack of byproducts found in the elution samples which renders the resultant purified protein or peptide more suitable for a therapeutic use.

Another aspect of the present invention comprises a purified or isolated RBP-7 polypeptide or a fragment or a variant thereof.

In a preferred embodiment, the RBP-7 polypeptide comprises an amino acid sequence of SEQ ID No. 29 or a fragment or a variant thereof. In a further embodiment, the present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 29.

The RBP-7 polypeptide of the amino acid sequence of SEQ ID No. 29 has 1312 amino acids in length. This 1312 amino acid sequence harbors notably potential sites indicating post-translational modifications such as 8 N-glycosylation sites, 72 phosphorylation sites, 8 N-myristoylation sites and 4 amidation sites. The location of these sites is referred to in the appended Sequence Listing when disclosing the features of the amino acid sequence of SEQ ID No. 29.

The RBP-7 polypeptide shares some homology in amino acid sequence with another retinoblastoma binding protein, namely human RBP-1 (Fattaey et al., 1993). More precisely, a 48% identity has been found between RBP-7 and RBP-1 for the amino acid sequence beginning at position 1 and ending at position 790 of RBP-7. A 30% identity has been found for the amino acid sequence beginning at position 791 and ending at position 1312 of RBP-7.

A further object of the present invention concerns a purified or isolated polypeptide which is encoded by a nucleic acid comprising a nucleotide sequence selected from the group

consisting of SEQ ID Nos 1, 4 and 5-28 or fragments or variants thereof. Preferably, the purified or isolated polypeptide comprises at least 10, at least 15, at least 20 or at least 25 consecutive amino acids of the polypeptides encoded by SEQ ID Nos 1, 4 and 5-28.

The invention includes a nucleic acid encoding a RBP-7 polypeptide comprising at least one of the biallelic markers of the present invention, more particularly at least one of the biallelic markers defined in SEQ ID No. 30-71.

More generally, the invention also pertains to a variant RBP-7 polypeptide comprising at least one amino acid substitution, addition or deletion, when compared with the sequence of SEQ ID No. 29. More particularly, the invention encompasses a RBP-7 protein or a fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 29 comprising at least one of the following amino acids:

- a Glycine residue at the amino acid position 293 of SEQ ID No. 29;
- a Glutamic acid at the amino acid in position 963 of SEQ ID No. 29;
- a Methionine residue at the amino acid position 969 of SEQ ID No. 29.

A variant or mutated RBP-7 polypeptide comprises amino acid changes of at least one amino acid substitution, deletion or addition, preferably from 1 to 10, 20 or 30 amino acid substitutions or additions. The amino acid substitutions are generally non conservative in terms of polarity, charge, hydrophilicity properties of the substitute amino acid when compared with the native amino acid. The amino acid changes occurring in such a mutated RBP-7 polypeptide may be determinant for the biological activity or for the capacity of the mutated RBP-7 polyeptide to be recognized by antibodies raised against a native RBP-7.

Such a variant or mutated RBP-7 protein may be the target of diagnostic tools, such as specific monoclonal or polyclonal antibodies, useful for detecting the mutated RBP-7 protein in a sample.

Are also part of the present invention polypeptides that are homologous to a RBP-7 polypeptide, especially a polypeptide of SEQ ID No. 29, or their fragments or variants.

The invention also encompasses a RBP-7 polypeptide or a fragment or a variant thereof in which at least one peptide bound has been modified as described in "Definitions".

The polypeptides according to the invention may also be prepared by the conventional methods of chemical synthesis, either in a homogenous solution or in solid phase. As an illustrative embodiment of such chemical polypeptide synthesis techniques, it may be cited the homogenous solution technique described by Houbenweyl in 1974.

The RBP-7 polypeptide, or a fragment or a variant thereof may thus be prepared by chemical synthesis in liquid or solid phase by successive couplings of the different amino acid

residues to be incorporated (from the N-terminal end to the C-terminal end in liquid phase, or from the C-terminal end to the N-terminal end in solid phase) wherein the N-terminal ends and the reactive side chains are previously blocked by conventional groups.

For solid phase synthesis the technique described by Merrifield (1965) may be used in particular.

ANTIBODIES

The polypeptides according to the present invention, especially the polypeptides of SEQ ID No. 29 are allowing the preparation of polyclonal or monoclonal antibodies that recognize the polypeptides of SEQ ID No. 29 or fragments thereof.

The antibodies may be prepared from hybridomas according to the technique described by Kohler and Milstein in 1975. The polyclonal antibodies may be prepared by immunization of a mammal, especially a mouse or a rabbit, with a polypeptide according to the invention that is combined with an adjuvant of immunity, and then by purifying of the specific antibodies contained in the serum of the immunized animal on a affinity chromatography column on which has previously been immobilized the polypeptide that has been used as the antigen.

The invention also concerns a purified or isolated antibody capable of specifically binding to the RBP-7 protein, more particularly to selected peptide fragments thereof, and more preferably polypeptides encoded by nucleic acids comprising one or more biallelic markers of the invention, or a variant thereof. In addition, the invention comprises antibodies capable of specifically binding to a fragment or variant of such a RBP-7 protein comprising an epitope of the RBP-7 protein, preferably an antibody capable of binding to a polypeptide comprising at least 10 consecutive amino acids, at least 15 consecutive amino acids, at least 20 consecutive amino acids, or at least 40 consecutive amino acids of a RBP-7 protein, more preferably an antibody capable of binding specifically to a variant or mutated RBP-7 protein or a fragment thereof and distinguishing between either two variants of RBP-7 or mutated RBP-7 and non-mutated RBP-7 protein.

The proteins expressed from a *RBP-7* DNA comprising at least one of the nucleic sequences of SEQ ID Nos 30-71 or a fragment or a variant thereof, preferably the nucleic sequences of the biallelic markers leading to an amino acid substitution, may also be used to generate antibodies capable of specifically binding to the expressed RBP-7 protein or fragments or variants thereof.

In another embodiment, polyclonal or monoclonal antibodies according to the invention are raised against a RBP-7 polypeptide comprising at least one of the following amino acids:

- a Glycine residue at the amino acid position 293 of SEQ ID No. 29;
- a Glutamic acid at the amino acid in position 963 of SEQ ID No. 29;

- a Methionine residue at the amino acid position 969 of SEQ ID No. 29.

Alternatively, the antibodies may be capable of binding fragments of the RBP-7 protein which comprise at least 10 amino acids encoded by the sequences of SEQ ID Nos 1 and 4, preferably comprising at least one of the sequences of SEQ ID Nos 30-71 or a fragment or a variant thereof. In some embodiments, the antibodies may be capable of binding fragments of the RBP-7 protein which comprise at least 15 amino acids encoded by the sequences of SEQ ID Nos 1 and 4, preferably comprising at least one of the sequences of SEQ ID Nos 30-71 or a fragment or a variant thereof. In other embodiments, the antibodies may be capable of binding fragments of the RBP-7 protein which comprise at least 25 amino acids encoded by the sequences of SEQ ID Nos 1 and 4, preferably comprising at least one of the sequences of SEQ ID Nos 30-71 or a fragment or a variant thereof. In further embodiments, the antibodies may be capable of binding fragments of the RBP-7 protein which comprise at least 40 amino acids encoded by the sequences of SEQ ID Nos 1 and 4, preferably comprising at least one of the sequences of SEQ ID Nos 30-71 or a fragment or a variant thereof.

Both monoclonal antibodies and polyclonal antibodies are within the scope of the present invention. Monoclonal or polyclonal antibodies to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the RBP-7 protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler and Milstein, (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the RBP-7 protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the RBP-7 protein or a portion thereof can be prepared by immunizing suitable animals with the RBP-7 protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. Effective

polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al. (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum. Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

Consequently, the invention is also directed to a method for detecting specifically the presence of a polypeptide according to the invention in a biological sample, said method comprising the following steps :

a) bringing into contact the biological sample with an antibody according to the invention;

b) detecting the antigen-antibody complex formed.

Another aspect of the invention is a diagnostic kit for in vitro detecting the presence of a polypeptide according to the present invention in a biological sample, wherein said kit comprises:

a) a polyclonal or monoclonal antibody as described above, optionally labeled;

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the above-mentioned monoclonal or polyclonal antibody is not labeled by itself.

METHODS FOR SCREENING SUBSTANCES INTERACTING WITH A RBP-7 POLYPEPTIDE

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to the RBP-7

protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for RBP-7 or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of the RBP-7 protein is brought into contact with the purified RBP-7 protein, for example the purified recombinant RBP-7 protein produced by a recombinant cell host as described hereinbefore, in order to form a complex between the RBP-7 protein and the putative ligand molecule to be tested.

The present invention pertains to methods for screening substances of interest that interact with a RBP-7 protein or one fragment or variant thereof. By their capacity to bind covalently or non-covalently to a RBP-7 protein or to a fragment or variant thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*.

In vitro, said interacting molecules may be used as detection means in order to identify the presence of a RBP-7 protein in a sample, preferably a biological sample.

A method for the screening of a candidate substance comprises the following steps :

a) providing a polypeptide comprising, consisting essentially of, or consisting of a RBP-7 protein or a fragment or a variant thereof;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance;

d) detecting the complexes formed between said polypeptide and said candidate substance.

In one embodiment of the screening method defined above, the complexes formed between the polypeptide and the candidate substance are further incubated in the presence of a polyclonal or a monoclonal antibody that specifically binds to the RBP-7 protein or to said fragment or variant thereof.

Various candidate substances or molecules can be assayed for interaction with a RBP-7 polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule comprises a polypeptide, this polypeptide may be the resulting expression product of a phage clone belonging to a phage-based random peptide library, or alternatively the polypeptide may be the resulting expression product of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a RBP-7 polypeptide or a fragment or a variant thereof, and optionally means useful to detect the complex formed between the RBP-7 polypeptide or its fragment or variant and the candidate substance. In a preferred embodiment

the detection means are monoclonal or polyclonal antibodies directed against the RBP-7 polypeptide or a fragment or a variant thereof.

A. Candidate Ligands Obtained Form Random Peptide Libraries

In a particular embodiment of the screening method, the putative ligand is the expression product of a DNA insert contained in a phage vector (Parmley and Smith, 1988). Specifically, random peptide phages libraries are used. The random DNA inserts encode for peptides of 8 to 20 amino acids in length (Oldenburg K.R. et al., 1992.; Valadon P., et al., 1996.; Lucas A.H., 1994; Westerink M.A.J., 1995; Castagnoli L. et al. (Felici F.), 1991). According to this particular embodiment, the recombinant phages expressing a protein that binds to the immobilized RBP-7 protein is retained and the complex formed between the RBP-7 protein and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the RBP-7 protein.

Once the ligand library in recombinant phages has been constructed, the phage population is brought into contact with the immobilized RBP-7 protein. Then the preparation of complexes is washed in order to remove the non-specifically bound recombinant phages. The phages that bind specifically to the RBP-7 protein are then eluted by a buffer (acid pH) or immunoprecipitated by the monoclonal antibody produced by the hybridoma anti-RBP-7, and this phage population is subsequently amplified by an over-infection of bacteria (for example E. coli). The selection step may be repeated several times, preferably 2-4 times, in order to select the more specific recombinant phage clones. The last step involves characterizing the peptide produced by the selected recombinant phage clones either by expression in infected bacteria and isolation, expressing the phage insert in another host-vector system, or sequencing the insert contained in the selected recombinant phages.

B. Candidate Ligands Obtained Through A Two-Hybrid Screening Assay

The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields and Song, 1989), and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in the US Patent No. US 5,667,973 and the US Patent No. 5,283,173 (Fields et al.) the technical teachings of both patents being herein incorporated by reference.

The general procedure of library screening by the two-hybrid assay may be performed as described by Harper et al. (Harper JW et al., 1993) or as described by Cho et al. (1998) or also Fromont-Racine et al. (1997).

The bait protein or polypeptide comprises, consists essentially of, or consists of a RBP-7 polypeptide or a fragment or variant thereof.

More precisely, the nucleotide sequence encoding the RBP-7 polypeptide or a fragment or variant thereof is fused to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example pAS2 or pM3.

Then, a human cDNA library is constructed in a specially designed vector, such that the human cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcriptional domain of the GAL4 protein. Preferably, the vector used is the pACT vector. The polypeptides encoded by the nucleotide inserts of the human cDNA library are termed "pray" polypeptides.

A third vector contains a detectable marker gene, such as β galactosidase gene or CAT gene that is placed under the control of a regulation sequence that is responsive to the binding of a complete Gal4 protein containing both the transcriptional activation domain and the DNA binding domain. For example, the vector pG5EC may be used.

Two different yeast strains are also used. As an illustrative but non limiting example the two different yeast strains may be the followings :

Y190, the phenotype of which is (*MATa, Leu2-3, 112 ura3-12, trp1-901, his3-D200, ade2-101, gal4Dgal180D URA3 GAL-LacZ, LYS GAL-HIS3, cyh*^{r});

Y187, the phenotype of which is (*MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112 URA3 GAL-lacZmet*$^-$), which is the opposite mating type of Y190.

Briefly, 20 μg of pAS2/RBP-7 and 20 μg of pACT-cDNA library are co-transformed into yeast strain Y190. The transformants are selected for growth on minimal media lacking histidine, leucine and tryptophan, but containing the histidine stnthesis inhibitor 3-AT (50 mM). Positive colonies are screened for beta galactosidase by filter lift assay. The double positive colonies (*His*$^+$, *β-gal*$^+$) are then grown on plates lacking histidine, leucine, but containing tryptophan and cycloheximide (10 mg/ml) to select for loss of pAS2/RBP-7 plasmids bu retention of pACT-cDNA library plasmids. The resulting Y190 strains are mated with Y187 strains expressing RBP-7 or non-related control proteins; such as cyclophilin B, lamin, or SNF1, as *Gal4* fusions as described by Harper et al. (Harper JW et al., 1993) and by Bram et al. (Bram RJ et al., 1993), and screened for β galactosidase by filter lift assay. Yeast clones that are *β gal-* after mating with the control *Gal4* fusions are considered false positives.

In another embodiment of the two-hybrid method according to the invention, interaction between RBP-7 or a fragment or variant thereof with cellular proteins may be assessed using the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech), the disclosure of which is incorporated herein by reference, nucleic acids encoding the RBP-7

protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. A desired cDNA, preferably human cDNA, is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain interaction between RBP-7 and the protein or peptide encoded by the initially selected cDNA insert.

C. Candidate Ligand Obtained Through Biosensor Assay

Proteins interacting with the RBP-7 protein or portions thereof can also be screened by using an Optical Biosensor as described in Edwards et Leatherbarrow (1997), the disclosure of which is incorporated herein by reference. The main advantage of the method is that it allows the determination of the association rate between the protein and other interacting molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethl dextran matrix) and a sample of test molecules is placed in contact with the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it occurs in the evanescent field (which extend a few hundred nanometers from the sensor surface). In these screening assays, the target molecule can be the RBP-7 protein or a portion thereof and the test sample can be a collection of proteins extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or chemical libraries,or phage displayed peptides. The tissues or cells from which the test proteins are extracted can originate from any species.

METHOD FOR SCREENING LIGANDS THAT MODULATE THE EXPRESSION OF THE *RBP-7* GENE

The present invention also concerns a method for screening substances or molecules that are able to increase, or in contrast to decrease, the level of expression of the *RBP-7* gene. Such a method may allow the one skilled in the art to select substances exerting a regulating effect on the expression level of the *RBP-7* gene and which may be useful as active ingredients included in pharmaceutical compositions for treating patients suffering from deficiencies in the regulation of expression of the *RBP-7* gene.

Thus, another aspect of the present invention is a method for the screening of a candidate substance or molecule, said method comprising the following steps :

a) providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID Nos : 1, 4, 30-75 or the sequences complementary thereto or a fragment or a variant thereof;

b) obtaining a candidate substance, and

c) determining the ability of the candidate substance to modulate the expression levels of the nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID No: 1, 4, 30-75 or the sequences complementary thereto or a fragment or a variant thereof.

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a recombinant vector that allows the expression of a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID No: 1, 4, 30-75 or the sequences complementary thereto or a fragment or a variant thereof or, alternatively, the kit may comprise a recombinant cell host containing such recombinant vectors.

Another subject of the present invention is a method for screening molecules that modulate the expression of the RBP-7 protein. Such a screening method comprises the steps of :

a) cultivating a prokaryotic or an eukaryotic cell that has been transfected with a nucleotide sequence encoding the RBP-7 protein, placed under the control of its own promoter;

b) bringing into contact the cultivated cell with a molecule to be tested;

c) quantifying the expression of the RBP-7 protein.

In another embodiment of a method for screening of a candidate substance or molecule that modulates the expression of the *RBP-7* gene, the method comprises the following steps :

a) providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid comprises the nucleotide sequence of SEQ ID No. 2, the sequence complementary thereto, or à biologically active fragment or variant thereof located upstream a polynucleotide encoding a detectable protein;

b) obtaining a candidate substance, and

c) determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

Among the preferred polynucleotides encoding a detectable protein, there may be cited polynucleotides encoding β galactosidase, green fluorescent protein (GFP) and chloramphenicol acetyl transferase (CAT).

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a recombinant vector that allows the expression of a nucleotide sequence of SEQ ID No. 2 or a biologically active fragment or variant thereof located upstream a polynucleotide encoding a detectable protein.

For the design of suitable recombinant vectors useful for performing the screening methods described above, it will be referred to the section of the present specification wherein the preferred recombinant vectors of the invention are detailed.

Using DNA recombination techniques well known by the one skill in the art, the RBP-7 protein encoding DNA sequence is inserted into an expression vector, downstream from its promoter sequence. As an illustrative example, the promoter sequence of the *RBP-7* gene is contained in the nucleic acid of SEQ ID No. 2.

The quantification of the expression of the RBP-7 protein may be realized either at the mRNA level or at the protein level. In the latter case, polyclonal or monoclonal antibodies may be used to quantify the amounts of the RBP-7 protein that have been produced, for example in an ELISA or a RIA assay.

In a preferred embodiment, the quantification of the RBP-7 mRNA is realized by a quantitative PCR amplification of the cDNA obtained by a reverse transcription of the total mRNA of the cultivated RBP-7-transfected host cell, using a pair of primers specific for *RBP-7*.

Expression levels and patterns of *RBP-7* may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are incorporated herein by reference. Briefly, the *RBP-7* cDNA or the *RBP-7* genomic DNA described above, or fragments thereof, is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the *RBP-7* insert comprises at least 100 or more consecutive nucleotides of the genomic DNA sequence or the cDNA sequences, particularly those comprising at least one of SEQ ID Nos 30-71 or those encoding mutated RBP-7. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

METHODS FOR INHIBITING THE EXPRESSION OF A *RBP-7* GENE

Other therapeutic compositions according to the present invention comprise advantageously an oligonucleotide fragment of the nucleic sequence of *RBP-7* as an antisense tool or a triple helix tool that inhibits the expression of the corresponding *RBP-7* gene.

A- Antisense Approach

Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al. (Sczakiel G. et al., 1995).

Preferably, the antisense tools are choosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of the *RBP-7* mRNA. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targetted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of *RBP-7* that contains the translation initiation codon ATG.

The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They comprise a nucleotide sequence complementary to the targeted sequence of the *RBP-7* genomic DNA, the sequence of which can be determined using one of the detection methods of the present invention. In a preferred embodiment, the antisense oligonucleotide are able to hybridize with at least one of the splicing sites of the targeted *RBP-7* gene, or with the 3'UTR of the 5'UTR. The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the *RBP-7* mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green et al., (1986) and Izant and Weintraub, (1984), the disclosures of which are incorporated herein by reference.

In some strategies, antisense molecules are obtained by reversing the orientation of the *RBP-7* coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of *RBP-7* antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, suitable antisense strategies are those described by Rossi et al. (1991), in the International Applications Nos. WO 94/23026, WO 95/04141, WO 92/18522 and in the European Patent Application No. EP 0 572 287 A2, the disclosures of which are incorporated herein by reference in their entireties.

An alternative to the antisense technology that is used according to the present invention involves using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme

involves (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present invention are prepared as described by Sczakiel et al. (1995), the specific preparation procedures being referred to in said article being herein incorporated by reference.

B- Triple Helix Approach

The *RBP-7* genomic DNA may also be used to inhibit the expression of the *RBP-7*gene based on intracellular triple helix formation.

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene.

Similarly, a portion of the *RBP-7* genomic DNA can be used to study the effect of inhibiting *RBP-7* transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the *RBP-7* genomic DNA are contemplated within the scope of this invention.

To carry out gene therapy strategies using the triple helix approach, the sequences of the *RBP-7* genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting *RBP-7* expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting *RBP-7* expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the *RBP-7* gene.

The oligonucleotides can be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced *RBP-7* expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the *RBP-7* gene in cells which have been treated with the oligonucleotide.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced in vivo using the techniques described above in the antisense approach at a dosage calculated based on the in vitro results, as described in antisense approach.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3′ end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al. (1989), which is hereby incorporated by this reference.

Throughout this application, various references are referred to within parentheses. The disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the sate of the art to which this invention pertains.

EXAMPLES

Example 1

Analysis Of The mRNAs Encoding A RBP-7 Polypeptide Synthesized By The Cells.

RBP-7 cDNA was obtained as follows : 4µl of ethanol suspension containing 1 mg of human prostate total RNA (Clontech laboratories, Inc., Palo Alto, USA; Catalogue N. 64038-1) was centrifuged, and the resulting pellet was air dried for 30 minutes at room temperature.

First strand cDNA synthesis was performed using the AdvantageTM RT-for- PCR kit (Clontech laboratories Inc., catalogue N. K1402-1). 1 µl of 20 mM solution of a specific oligo dT primer was added to 12.5 µl of RNA solution in water, heated at 74°C for 2.5 min and rapidly quenched in an ice bath. 10 µl of 5 x RT buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 2.5 µl of dNTP mix (10 mM each), 1.25 µl of human recombinant placental RNA inhibitor were mixed with 1 ml of MMLV reverse transcriptase (200 units). 6.5 µl of this solution were added to RNA-primer mix and incubated at 42°C for one hour. 80 µl of water were added and the solution was incubated at 94°C for 5 minutes.

5µl of the resulting solution were used in a Long Range PCR reaction with hot start, in 50 µl final volume, using 2 units of rTHXL, 20 pmol/µl of each of 5'-CCCTTGATGAGCCTCCCTATTTGACAG-3' (SEQ ID No. 137) and 5'-CGCATTGAAATTCCCACGTCGTATTGCCAG-3' (SEQ ID No. 138) primers with 35 cycles of elongation for 6 minutes at 67°C in thermocycler.

The amplification products corresponding to both cDNA strands are partially sequenced in order to ensure the specificity of the amplification reaction.

Results of Nothern blot analysis of prostate mRNAs support the existence of a major RBP-7 cDNA having about 6 kb in length, which is approximately the size of the longest possible RBP-7 transcript.

EXAMPLE 2

Detection Of *RBP-7* Biallelic Markers: DNA Extraction

Donors were unrelated and healthy. They presented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the detection of the biallelic markers.

30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume : 10 mM Tris pH7.6; 5 mM $MgCl_2$; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M
- 200 µl SDS 10%
- 500 µl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA).

To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below.

The pool was constituted by mixing equivalent quantities of DNA from each individual.

EXAMPLE 3

Detection Of The Biallelic Markers: Amplification Of Genomic DNA By PCR

The amplification of specific genomic sequences of the DNA samples of example 2 was carried out on the pool of DNA obtained previously. In addition, 50 individual samples were similarly amplified.

PCR assays were performed using the following protocol:

Final volume	25 µl
DNA	2 ng/µl
$MgCl_2$	2 mM
dNTP (each)	200 µM
primer (each)	2.9 ng/µl
Ampli Taq Gold DNA polymerase	0.05 unit/µl
PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl	1x

Each pair of primers was designed using the sequence information of the *RBP-7* gene disclosed herein and the OSP software (Hillier & Green, 1991). This pair of primers was about 20 nucleotides in length and had the sequences disclosed in Table 1 in the columns labeled PU and RP.

TABLE 1

Amplicon	Amplification primer PU SEQ ID No.	Amplification primer RP SEQ ID No.
5-124	72	87
5-127	73	88
5-128	74	89
5-129	75	90
5-130	76	91
5-131	77	92
5-133	78	93
5-135	79	94
5-136	80	95
5-140	81	96
5-143	82	97
5-145	83	98
5-148	84	99
99-1437	85	100
99-1442	86	101

Preferably, the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification which was useful for sequencing.

Primers PU contain the following additional PU 5' sequence : TGTAAAACGACGGCCAGT (SEQ ID No. 139); primers RP contain the following RP 5' sequence : CAGGAAACAGCTATGACC (SEQ ID No. 140).

The synthesis of these primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

DNA amplification was performed on a Genius II thermocycler. After heating at 95°C for 10 min, 40 cycles were performed. Each cycle comprised: 30 sec at 95°C, 54°C for 1 min, and 30 sec at 72°C. For final elongation, 10 min at 72°C ended the amplification. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

EXAMPLE 4

Detection Of The Biallelic Markers: Sequencing Of Amplified Genomic DNA And Identification Of Polymorphisms.

The sequencing of the amplified DNA obtained in example 3 was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis [ABI Prism DNA Sequencing Analysis software (2.1.2 version) and the above mentioned "Trace" basecaller].

The sequence data were further evaluated using the above mentioned polymorphism analysis software designed to detect the presence of biallelic markers among the pooled amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position as described previously.

Sixteen fragments of amplification were analyzed. In these segments, 21 biallelic markers were detected. The localization of the biallelic markers is as shown in Table 2.

TABLE 2

Amplicon	BM	Marker Name	Localization in *RBP-7*	BM position in SEQ ID No. 1	Polymor-phism	SEQ ID No.	
						Allele 1	Allele 2
5-124	A1	5-124-273	Intron 5	72794	A*/G	30	51
5-127	A2	5-127-261	Intron 8	88073	A/C*	31	52
5-128	A3	5-128-60	Intron 8	93714	Del(GT)	32	53
5-129	A4	5-129-144	Intron 9	97152	Del(T)	33	54
5-130	A5	5-130-257	Exon 11	99098	A*/G	34	55
5-130	A6	5-130-276	Exon 11	99117	A/G	35	56
5-131	A7	5-131-395	Intron 12	103806	A*/T	36	57
5-133	A8	5-133-375	Intron 14	106940	ins(A)	37	58
5-135	A9	5-135-155	Intron 15	108106	ins(A)	38	59
5-135	A10	5-135-198	Intron 15	108149	ins(GTTT)	39	60
5-135	A11	5-135-357	Intron 15	108308	A*/G	40	61
5-136	A12	5-136-174	Exon 16	108471	C/T*	41	62
5-140	A13	5-140-120	Intron 18	134134	C/T*	42	63
5-140	A14	5-140-348	Intron 19	134362	ins(A)	43	64
5-140	A15	5-140-361	Intron 19	134374	ins(CA)	44	65
5-143	A16	5-143-101	Exon 20	146345	A/C	45	66
5-143	A17	5-143-84	Exon 20	146328	A/G*	46	67
5-145	A18	5-145-24	Intron 20	150329	A*/G	47	68
5-148	A19	5-148-352	Exon 24	160031	G/T	48	69
99-1437	A20	99-1437-325	Intron 8	90842	A/G	49	70
99-1442	A21	99-1442-224	Intron 9	97122	G/T	50	71

* the most frequent allele in the tested Caucasian control population

EXAMPLE 5

Validation Of The Polymorphisms Through Microsequencing

The biallelic markers identified in example 4 were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out for each individual DNA sample described in Example 2.

Amplification from genomic DNA of individuals was performed by PCR as described above for the detection of the biallelic markers with the same set of PCR primers (Table 1).

The preferred primers used in microsequencing were about 23 nucleotides in length and hybridized just upstream of the considered polymorphic base. According to the invention, the primers used in microsequencing are detailed in Table 3.

TABLE 3

Marker Name	Mis. 1 in SEQ ID No.	Mis. 2 in SEQ ID No.
5-124-273	102	123
5-127-261	103	123
5-128-60	104	-
5-129-144	105	-
5-130-257	106	125
5-130-276	107	126
5-131-395	108	127
5-133-375	109	-
5-135-155	110	-
5-135-198	111	-
5-135-357	112	128
5-136-174	113	129
5-140-120	114	130
5-140-348	115	-
5-140-361	116	-
5-143-101	117	131
5-143-84	118	132
5-145-24	119	133
5-148-352	120	134
99-1437-325	121	135
99-1442-224	122	136

The microsequencing reaction was performed as follows :

After purification of the amplification products, the microsequencing reaction mixture was prepared by adding, in a 20µl final volume: 10 pmol microsequencing oligonucleotide, 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM $MgCl_2$), and the two appropriate fluorescent ddNTPs (Perkin Elmer, Dye Terminator Set 401095) complementary to the nucleotides at the polymorphic site of each biallelic marker tested, following the manufacturer's recommendations. After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The unincorporated dye terminators were then removed by ethanol precipitation. Samples were finally resuspended in formamide-EDTA loading buffer and heated for 2 min at 95°C before being loaded on a polyacrylamide sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Following gel analysis, data were automatically processed with software that allows the determination of the alleles of biallelic markers present in each amplified fragment.

The software evaluates such factors as whether the intensities of the signals resulting from the above microsequencing procedures are weak, normal, or saturated, or whether the signals are ambiguous. In addition, the software identifies significant peaks (according to shape and height criteria). Among the significant peaks, peaks corresponding to the targeted site are identified based on their position. When two significant peaks are detected for the same position, each sample is categorized classification as homozygous or heterozygous type based on the height ratio.

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims. All documents cited herein are incorporated herein by reference in their entirety.

REFERENCES :

The disclosures of all of the following publications are incorporated herein by reference in their entireties:

Abbondanzo SJ et al., 1993, Methods in Enzymology, Academic Press, New York, pp. 803-823.

Adra et al.,1987, Gene, **60**:65-74

Anton M. et al., 1995, J. Virol., **69** : 4600-4606.

Araki K. et al., 1995, Proc. Natl. Acad. Sci. USA, **92** : 160-164.

Baubonis et al., 1993, Nucleic Acids Res., **21** : 2025-2029.

Beard et al., 1980, Virology, Vol. 75:81

Beaucage et al., *Tetrahedron Lett* 1981, **22**: 1859-1862

Bolmont et al., J. of Submicroscopic cytology and pathology, 1990, **22**:117-122

Bradley A., 1987, Production and analysis of chimaeric mice. *In* : E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells : A practical approach. IRL Press, Oxford, pp.113.

Bram RJ et al., 1993, Mol. Cell Biol., **13** : 4760-4769

Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979;**68**:109-151

Castagnoli L. et al. (Felici F.), 1991, J. Mol. Biol., **222**:301-310

Chai H. et al., 1993, Biotechnol. Appl. Biochem., **18**:259-273

Cho RJ et al., 1998, Proc. Natl. Acad. Sci. USA, **95**(7) : 3752-3757.

Chou J.Y., 1989, Mol. Endocrinol., **3** : 1511-1514.

Compton J. *Nature*. 1991 Mar 7; **350**(6313): 91-92.

Current Protocols in Molecular Biology, 1989, Ausubel FM et al. (eds), Greene Publishing Associates, Sections 9.10-9.14.

Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2

Defeo-Jones et al. Nature, 1991, Vol. 352 : 251-254

Eckner R. et al., 1991, EMBO J., **10** : 3513-3522.

Edwards et Leatherbarrow, Analytical Biochemistry, 246, 1-6 (1997)

Feldman and Steg, 1996, Medecine/Sciences, synthese, **12**:47-55

Felgner et al., 1987, Proc. Natl. Acad. Sci., **84**:7413

Fields and Song, 1989, Nature, **340** : 245-246

Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Flotte et al., 1992, Am. J. Respir. Cell Mol. Biol., **7** : 349-356.

Fraley et al., 1980, J. Biol. Chem., **255**:10431).

Fromont-Racine M. et al., 1997, Nature Genetics, **16**(3) : 277-282.

Fuller S.A. et al., 1996, Immunology in Current Protocols in Molecular Biology, Ausubel et al. Eds, John Wiley & Sons, Inc., USA

Furth P.A. et al., 1994, Proc. Natl Acad. Sci USA, **91** : 9302-9306.

Gossen M. et al., 1992, Proc. Natl. Acad. Sci. USA, **89** : 5547-5551.

Gossen M. et al., 1995, Science, **268** : 1766-1769.

Graham, 1984, EMBO J., **3**:2917

Green et al., *Ann. Rev. Biochem.* **55**:569-597 (1986)

Griffin et al. *Science* **245**:967-971 (1989)

Grompe, M. et al., *Proc. Natl. Acad. Sci. U.S.A* 1989; **86**:5855-5892

Grompe, M. *Nature Genetics* 1993; **5**:111-117

Gu H. et al., 1993, Cell, **73** : 1155-1164.

Gu H. et al., 1994, Science, **265** : 103-106.

Guatelli JC et al., 1990, Proc. Natl. Acad. Sci. USA, **87** : 1874-1878.

Guzman, 1981, Cell, **23** : 175

Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS, *Nat Genet* 1996;**14**(4):441-447

Hames BD and Higgins SJ, 1985, "Nucleic acid hybridization : a practical approach", Hames and Higgins Ed., IRL Press, Oxford.

Harper JW et al., 1993, Cell, **75** : 805-816

Harris H. et al., 1969, Nature, **223** : 363-368

Helin et al., 1992, Cell, **70** : 337-350

Hillier L. and Green P. *Methods Appl.*, 1991, **1**: 124-8.

Hoess et al., 1986, Nucleic Acids Res., **14** : 2287-2300.

Houbenweyl, 1974, in Meuthode der Organischen Chemie, E. Wunsch Ed., Volume 15-I et 15-II, Thieme, Stuttgart

Huygen et al., 1996, Nature Medicine, **2**(8):893-898

Izant JG, Weintraub H, *Cell* 1984 Apr;**36**(4):1007-15

Julan et al., 1992, J. Gen. Virol., **73** : 3251 – 3255.

Kaneda et al., 1989, Science, **243**:375

Kanegae Y. et al., Nucl. Acids Res., **23** : 3816-3821.

Kim et al., 1994, Mol. And Cell Biol., **14**(11) : 7256-7264

Kim U. –J., et al., 1996, Genomics, **34** : 213-218.

Koch Y., 1977, Biochem. Biophys. Res. Commun., **74**:488-491

Kohler G. and Milstein C., 1975, Nature, **256** : 495.

Koller et al., 1992, Annu. Rev. Immunol., **10** : 705-30.

Kort et al., 1983, Nucleic Acids Research, **11**:8287-8301

Kozal MJ, Shah N, Shen N, Yang R, Fucini R, Merigan TC, Richman DD, Morris D, Hubbell E, Chee M, Gingeras TR, *Nat Med* 1996;**2**(7):753-759

Lenhard T. et al., 1996, Gene, **169**:187-190

Levrero et al., 1991, Gene, **101**:195

Linton M.F. et al., 1993, J. Clin. Invest., **92** : 3029-3037.

Liu Z. et al., 1994, Proc. Natl. Acad. Sci. USA, **91** : 4528-4262.

Lucas A.H., 1994, In : Development and Clinical Uses of Haempophilus b Conjugate;

Mansour S.L. et al., 1988, Nature, **336** : 348-352.

Mansour SL et al., 1988, Nature, **336** : 348-352.

Manz et al., *Adv in Chromatogr* 1993; **33**:1-66

Marshall R. L., et al., *PCR Methods and Applications* **4**: 80-84 (1994)

McCormick et al., 1994, Genet. Anal. Tech. Appl., **11** : 158-164.

McLaughlin et al., 1989, J. Virol., **62** : 1963 – 1973.

Merrifield RB, 1965, Nature, **207**(996): 522-523.

Merrifield RB., 1965, Science, **150**(693): 178-185.

Midoux, 1993, Nucleic Acids Research, **21**:871-878

Muzyczka et al., 1992, Cuur. Topics in Micro. and Immunol., **158** : 97-129.

Nada S. et al., 1993, Cell, **73** : 1125-1135.

Nagy A. et al., 1993, Proc. Natl. Acad. Sci. USA, **90** : 8424-8428.

Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;**68**:90-98

Neda et al., 1991, J. Biol. Chem., **266** : 14143 – 14146.

O'Reilly et al., 1992, Baculovirus expression vectors : a Laboratory Manual. W.H. Freeman and Co., New York Adra et al., 1987, Gene, **60**:65-74

Ohno et al., 1994, Sciences, **265**:781-784

Oldenburg K.R. et al., 1992, Proc. Natl. Acad. Sci., **89**:5393-5397.

Otterson et al., 1993, Oncogene, **8** : 949-957

Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973)

Pagano et al., 1967, J. Virol., **1**:891

Parmley and Smith, Gene, 1988, **73**:305-318

Pastore, 1994, Circulation, **90**:I-517

PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press

Pease S. ans William R.S., 1990, Exp. Cell. Res., **190** : 209-211.

Peterson et al., 1993, Proc. Natl. Acad. Sci. USA, **90** : 7593-7597.

Porath J et al., 1975, Nature, **258**(5536) : 598-599.

Reid L.H. et al., 1990, Proc. Natl. Acad. Sci. USA, **87** : 4299-4303.

Rettlez and Basenga, 1987, Mol. Cell. Biol., **7**:1676-1685

Robertson E., 1987, Embryo-derived stem cell lines. In : E.J. Robertson Ed. *Teratocarcinomas and embrionic stem cells : a practical approach.* IRL Press, Oxford, pp. 71.

Rossi et al., *Pharmacol. Ther.* **50**:245-254, (1991)

Roth J.A. et al., 1996, Nature Medicine, **2**(9):985-991

Rougeot, C. et al.,. *Eur. J. Biochem.* **219** (3): 765-773, 1994

Roux et al., 1989, Proc. Natl Acad. Sci. USA, **86** : 9079 – 9083.

Sakai et al., 1995, Genomics, **30** : 98-101

Sambrook, J. Fritsch, E. F., and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York

Samulski et al., 1989, J. Virol., **63** : 3822-3828.

Sanchez-Pescador R., 1988, J. Clin. Microbiol., **26**(10):1934-1938

Sandou et al., 1994, Science, **265** : 1875-1878.

Sauer B. et al., 1988, Proc. Natl. Acad. Sci. USA, **85** : 5166-5170.

Schedl A. et al., 1993a, Nature, **362** : 258-261.

Schedl et al., 1993b, Nucleic Acids Res., **21** : 4783-4787.

Sczakiel G. et al., 1995, Trends Microbiol., 1995, **3**(6):213-217

Shay J.W. et al., 1991, Biochem. Biophys. Acta, **1072** : 1-7.

Sheffield, V.C. et al, *Proc. Natl. Acad. Sci. U.S.A* 1991; **49**:699-706

Shizuya et al., 1992, Porc. Natl. Acad. Sci. USA **89** : 8794-8797.

Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW, *Nat Genet* 1996;**14**(4):450-456

Smith et al., 1983, Mol. Cell. Biol., **3**:2156-2165.

Sosnowski RG, Tu E, Butler WF, O'Connell JP, Heller MJ, *Proc Natl Acad Sci U S A* 1997;**94**(4):1119-1123

Sternberg N.L., 1992, Trends Genet., **8** : 1-16.

Sternberg N.L., 1994, Mamm. Genome, **5** : 397-404.

Tacson et al., 1996, Nature Medicine, **2**(8):888-892

Te Riele et al., 1990, Nature, **348** : 649-651.

Thomas K.R. et al., 1986, Cell, **44** : 419-428.

Thomas K.R. et al., 1987, Cell, **51** : 503-512.

Thomson et al., 1998, *Science* **282**, 1145-1147

Urdea M.S., 1988, Nucleic Acids Research, **11**: 4937-4957

Urdea MS et al., 1991, Nucleic Acids Symp Ser., **24**: 197-200.

Vaitukaitis J. et al., 1971, J. Clin. Endocrinol. Metab., **33** : 988-991

Valadon P., et al., 1996, J. Mol. Biol., **261**:11-22.

Van der Lugt et al., 1991, Gene, **105** : 263-267.

Vlasak R. et al., 1983, Eur. J. Biochem., **135**:123-126

Wabiko et al., 1986, DNA, **5**(4):305-314

Walker G. T. et al., *Clin. Chem.* **42**:9-13 [1996]

Westerink M.A.J., 1995, Proc. Natl. Acad. Sci., **92**:4021-4025

White, B.A. Molecular Cloning to Genetic Engineering Ed. in Methods in Molecular Biology **67**: Humana Press, Totowa 1997

White, M.B. et al., *Genomics* 1992; **12**:301-306

Wood S.A. et al., 1993, Proc. Natl. Acad. Sci. USA, **90** : 4582-4585.

Yagi T. et al., 1990, Proc. Natl. Acad. Sci. USA, **87** : 9918-9922.

Zhenlin et al., Gene, 1989, **78**:243-254

Zou Y.R. et al., 1994, Curr. Biol., **4** : 1099-1103.